



HAL
open science

Immunization of networks with non-overlapping community structure

Zakariya Ghalmane, Mohammed El Hassouni, Hocine Cherifi

► **To cite this version:**

Zakariya Ghalmane, Mohammed El Hassouni, Hocine Cherifi. Immunization of networks with non-overlapping community structure. *Social Network Analysis and Mining*, 2019, 9 (1), 10.1007/s13278-019-0591-9 . hal-02377044

HAL Id: hal-02377044

<https://hal.science/hal-02377044>

Submitted on 22 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Immunization of networks with non-overlapping community structure

Zakariya Ghalmane · Mohammed El
Hassouni · Hocine Cherifi

Received: date / Accepted: date

Abstract Although community structure is ubiquitous in complex networks, few works exploit this topological property to control epidemics. In this work, devoted to networks with non-overlapping community structure (i.e, a node belongs to a single community), we propose and investigate three global immunization strategies. In order to characterize the influence of a node, various pieces of information are used such as the number of communities that the node can reach in one hop, the nature of the links (intra community links, inter community links), the size of the communities, and the interconnection density between communities. Numerical simulations with the Susceptible-Infected-Removed (SIR) epidemiological model are conducted on both real-world and synthetic networks. Experimental results show that the proposed strategies are more effective than classical alternatives that are agnostic of the community structure. Additionally, they outperform alternative local and global strategies designed for modular networks.

Keywords Community structure · Immunization strategy · Epidemic spreading · Influence · Centrality · SIR model

Zakariya Ghalmane
LRIT URAC No 29, Faculty of Science, Rabat IT center, Mohammed V University, Rabat,
Morocco
E-mail: zakaria.ghalmane@gmail.com

Mohammed El Hassouni
LRIT URAC No 29, Faculty of Science, Rabat IT center, Mohammed V University, Rabat,
Morocco
E-mail: mohamed.elhassouni@gmail.com

Hocine Cherifi
LE2I UMR 6306 CNRS, University of Burgundy, Dijon, France
E-mail: hocine.cherifi@u-bourgogne.fr

1 Introduction

1 Epidemic outbreaks represent a tremendous threat to human life, since we live
2 in an ever more connected world. Immunization through vaccination is a solution
3 that protects individuals and prevents them from transmitting infectious diseases
4 to other people living in the same social group. However, immunizing every indi-
5 vidual in the society may prove impossible in cases where time or resources are
6 limited. To address this problem, immunization strategies are the essential tech-
7 niques to decrease the chances of epidemic outbreaks. It aims to immunize a few
8 key nodes to achieve effectively the goal of reducing or stopping the spread of
9 infectious diseases. Immunization strategies can be classified into two categories:
10 global or local immunization strategies. Global immunization strategies require
11 the knowledge of the entire network, hence their effectiveness. They consist of
12 ranking nodes according to a specific centrality measure like Degree or Between-
13 ness centrality. Nodes with high centrality measure are targeted for immunization.
14 Local immunization strategies are another group of immunization methods. They
15 are more or less agnostic about the topological structure of networks. In these
16 strategies, target nodes are found via local search. They require information only
17 at node level to find the targeted nodes for immunization.

18 The structure of networks is crucial in explaining epidemiological patterns.
19 In the past few years, many immunization strategies have been developed using
20 various topological properties of the network in order to mitigate and control the
21 epidemic outbreaks. Despite the fact that there is clear evidence that many social
22 networks show marked patterns of strong community structure [1,2,3,4], this prop-
23 erty needs more consideration. A network with a strong community structure con-
24 sists on cohesive subgroups of vertices that share many connections with members
25 of their group and few connections with vertices outside their group. Bridge nodes
26 are the ones that link different communities. They create a pathway of spread-
27 ing disease outside of their community. Their influence on epidemic spreading has
28 been particularly investigated in previous works [5,6,7,8]. Indeed, immunization
29 of these nodes allows confining the disease into the community where it starts.
30 However, one must not neglect the importance of the highly connected nodes em-
31 bedded into their community on the epidemic spreading process. In real-world
32 networks, the community structure strength can range from strong community
33 structure (few inter-community links) to weak community structure (high pro-
34 portion of inter-community links). The immunization strategies proposed in the
35 previous studies aim at targeting the key spreaders in networks with community
36 structure. However, either they do not exploit the community structure strength
37 or they do not use it properly. That is the reason why they are not suitable for
38 all types of networks. To solve this issue, we propose to make better use of the
39 information about the community structure in order to develop new immunization
40 strategies. The three immunization strategies presented in this work are intended
41 for various types of networks with community structure strength ranging from
42 well-defined to non-cohesive community structure. Our aim, therefore, is to relate
43 the impact of the community structure strength to the choice of an appropri-
44 ate immunization strategy. Additionally, our goal is to show that engaging more
45 topological properties of the community may enhance also the performance of the
46 immunization strategies.

In this work, we restrict our attention to networks where each node belongs to a single community. We also use a global approach. In other words, for each node of the network, an influence measure is computed and the nodes are ranked and immunized according to this measure. Thus, we propose and evaluate three methods:

i) The first proposed method targets nodes having a big inter-community influence. It is measured by the number of neighboring communities linked to the node.

ii) The second immunization method targets nodes which could have at the same time a high influence inside and outside their communities. Greater importance is given to those belonging to large communities since they could affect more nodes. This strategy is based on a weighted combination of the number of intra-community and inter-community links of each node in the network.

iii) The third method has the same objectives as the previous one. It is designed in order to take also into account the density variation of the communities.

The remainder of this paper is organized as follows. We preview the necessary background in Section 2. In Section 3, related works and immunization strategies are introduced. In Section 4, the proposed community-based strategies are defined. Section 5 introduces the experimental setting used in this work. In Section 6, the experimental results are presented. Finally, section 7 serves as a conclusion to the paper.

2 Background

In this section, we recall the definition of the immunization strategies that are used to mitigate an epidemic outbreak. In addition, we present the model used to simulate the epidemic spreading process in order to evaluate the performances of the different methods in the context of transmission dynamics. At least, a short outline of the community detection studies is presented.

2.1 Epidemiological model

The susceptible-infected (SI) and susceptible-infected-removed (SIR) models are widely used for infection dissemination and information diffusion in different fields. In this paper, we employ the SIR model to estimate the spreading capabilities of the nodes.

The SI model [9] is considered as the simplest form of all epidemic models. In this model, a node has only two possible states: a susceptible (S) or a infected (I) state. The model can be represented by the compartment diagram shown in Figure 1 (a). At first, all nodes are set to the susceptible state (individuals are with no immunity). After that, the state of a small proportion of nodes selected by a given immunization strategy is set to the infected state. At each time step, an infected node can infect its susceptible neighbors with the transmission rate λ . This process ends when there is no susceptible node in the network.

The Susceptible-Infected-Recovered (SIR) epidemic model [10,11] is used to simulate the spreading process in networks. In this model, there are three states for each node: susceptible (S), infected (I) and recovered (R). The infection mechanism of the SIR model is shown in Figure 1 (b). Initially, targeted nodes are

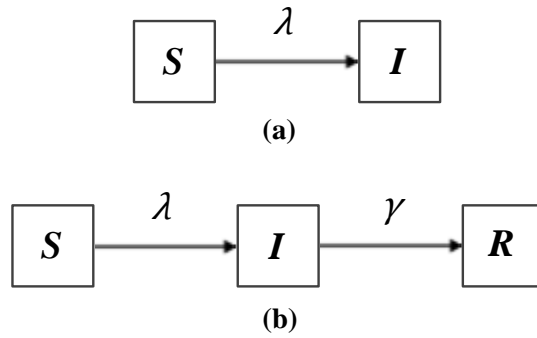


Fig. 1 The infection mechanism of the classic (a) SI model (b) SIR model.

90 chosen according to a given immunization strategy until a desired immunization
 91 coverage of the population is achieved, and their state is set to resistant R. All
 92 remaining nodes are in S state. After this initial set-up, infection starts from a
 93 random susceptible node. Its state changes to I. At each time step, the epidemic
 94 spreads from one infected node to a neighboring susceptible node according to the
 95 transmission rate of infection λ . Furthermore, infected nodes recover at rate γ , i.e.
 96 the probability of recovery of an infected node per time step is γ . If recovery oc-
 97 curs, the state of the recovered node is set from infected to resistant. The epidemic
 98 spreading process ends when there is no infected node in the network. After each
 99 simulation, we record the total number of recovered nodes (the epidemic size).

100 2.2 Immunization strategy

101 The goal of an immunization strategy [12] is to reveal the set of the most influ-
 102 ential spreaders in a given network. According to the amount of information they
 103 require about the overall structure of the network, they can be classified into two
 104 categories: Global and Local strategies. The first type of strategies requires infor-
 105 mation of the whole network topology, while the second group of strategies needs
 106 only the knowledge of network structure at node level.

107 Global immunization strategies are based on an ordering of all the nodes in
 108 the network in order to immunize them according to their rank. To do so, a so-
 109 called centrality measure is computed for each node of the network. It quantifies
 110 its ability to disseminate the disease inside the network. Degree and Betweenness
 111 are the most commonly used centrality measures to rank the nodes. Nodes are
 112 then targeted in the decreasing order of their rank from most central to less cen-
 113 tral node. Since all the nodes are involved in this process, the knowledge of the
 114 entire network is then required for these strategies. Local immunization strategies
 115 on the other hand are agnostic about the global structure of the network. They
 116 can operate with a very limited amount of information about a node. The most
 117 straightforward local strategy is uniform immunization that targets nodes in a to-
 118 tally random way without any information. Acquaintance [13] is another popular
 119 local immunization strategy which selects random neighbors of randomly selected
 120 nodes and immunizes them if they have been selected n times. Usually, global

121 strategies perform better than local strategies since they can use more informa-
122 tion about the topological properties of networks. However, the local strategies
123 are usually computationally more efficient.

124 2.3 Community structure

125 Many real-world networks exhibit a community structure, i.e., their nodes are
126 organized into modules, called communities. The first definitions of the commu-
127 nity structure were proposed by the Social network analysts. They studied the
128 structure of subgraphs. The Clique is the most prevalent concept [14]. A clique
129 is a complete subgraph such that everyone of its nodes is associated with all the
130 others. In general, communities are not complete graphs. In addition, in a clique
131 all nodes have identical role, while some nodes are more important than others
132 in communities, due to their heterogeneous linking patterns. Thus, this notion
133 cannot be viewed as an appropriate candidate for community definition. A very
134 widespread informal definition of the community concept considers it as a densely
135 interconnected group of vertices compared to the other vertices [15]. A community
136 is then a cohesive subset of nodes sparsely connected with the rest of the net-
137 work. This view has been challenged, recent works [16, 17, 18, 19] has shown that
138 communities may overlap as well. Some of the vertices can be shared by several
139 communities. In social networks for instance, individuals can take part to different
140 groups at the same time, such as work colleagues, friends or family.

141 Identifying the communities in networks may offer a clear idea on how the
142 network is organized. We can actually distinguish between nodes that are totally
143 embedded inside their groups and nodes that are located at the boundary of the
144 groups. These nodes may act as brokers between the communities of the networks
145 and could play a major role in the dynamics of spreading processes across the
146 network. Community detection in networks, also called network partitioning or
147 clustering is a not well characterized problem. Formal definitions may differ in the
148 way they consider these aspects of cohesion and separation of communities. There
149 is therefore no universal definition of the modules that one should be looking for.
150 Such ambiguity leaves a lot of freedom to propose various community detection
151 algorithms implementing differently the notion of community structure. In this
152 section, we present a representative set of methods and classify them according to
153 the approach they apply to uncover the communities.

154 2.3.1 Modularity based algorithms

155 Modularity is a widespread measure introduced by Newman and Girvan [20, 21],
156 which measures the quality of a community structure. It assesses the internal
157 connectivity of the identified communities through the number of intra- and inter-
158 community links. Modularity optimization based algorithms tend to identify the
159 best community structure in terms of modularity.

160 **FastGreedy** [22] is based on a greedy optimization approach. It starts with a
161 state in which each node constitutes its own community. The algorithm repeatedly
162 merges pairs of communities together to obtain larger ones. At each step, the joined
163 communities are selected by considering the largest increase (or smallest decrease)

164 in modularity. FastGreedy produces a hierarchy of community structures. The best
165 one is the one obtaining the maximal modularity.

166 **Louvain** proposed by Blondel *et al.* [23] is another optimization algorithm.
167 It relies on an improvement greedy optimization process. It includes a additional
168 agglomerative phase to improve the optimization approach. Initially as for Fast-
169 Greedy, each node constitutes its own community. After that, a greedy optimiza-
170 tion algorithm is applied to identify the communities. The second step consists
171 on forming a new network, where nodes represent the communities found during
172 the first phase. The inter-community links are aggregated and represented as links
173 between the new nodes, while the intra-community links are represented by self-
174 loops. The first phase is repeated to the new network, and the process ends when
175 stable communities are reached.

176 2.3.2 Random-walk based algorithms

177 Various algorithms utilize random walks in different ways in order to identify com-
178 munities in networks. In this work, we have retained one of the most influential
179 algorithm from this class.

180 **WalkTrap** [24] uses a hierarchical agglomerative clustering approach as for Fast-
181 Greedy, but with a different fusion criterion. It uses a distance measure based on
182 random walks. This algorithm is based on the idea that random walks tend to
183 get trapped into a community. If two nodes i and j are in the same community,
184 the probability to get to a third node k located in the same community through
185 a random walk should not be very different for both of them. The distance is
186 constructed by summing these differences over all nodes, with a correction for
187 degree.

188 2.3.3 Information based algorithms

189 The goal of these algorithms is to use the community structure so as to represent
190 the network using less information than that interpreted by the full adjacency
191 matrix. We retained two algorithms from this class.

192 **InfoMod** was proposed by Rosvall *et al.* [25], which uses a community matrix
193 and a membership vector as simplified representation of the network focusing
194 on the community structure. The first one is an adjacency matrix representing
195 communities instead of nodes, while the second one is a vector associating each
196 node to a community. This algorithm uses the mutual information measure in order
197 to measure the quantity of information from the original network contained in the
198 simplified representation. The best assignment among all possible assignment of
199 nodes to communities is the one associated with the maximal mutual information.

200 **InfoMap** [26] is another algorithm proposed by the same authors. It tends to
201 find the set of nodes (named communities) containing high intra-module informa-
202 tion flow and low inter-module information flow. The InfoMap algorithm is based
203 on a map equation. It is based on the information flow used to find a compressed
204 representation of a set of random walks through a graph. The partitions with high
205 quality are found by minimizing the quantity of information needed to represent
206 some random walk in the network. Indeed, the walker will probably stay longer
207 inside communities in a partition containing few inter-community links.

3 Related work

Immunization strategies aim to immunize a few key nodes to effectively achieve the goal of reducing or stopping the spread of infectious diseases. They can be classified as local or global strategies. In local immunization strategies, local information about randomly selected nodes is used in order to identify target nodes. As they do not need any information about the full network structure, they can be used in situations where it is unavailable. In global strategies, for each node of the network, one compute a measure of influence using local or global measure. Nodes are then ranked and immunized according to their influence value. Researchers have begun to pay more attention to the community structure in terms of epidemic dynamics [16, 27, 28, 29, 30]. Local and global strategies based on the community structure characteristics have been proposed. They can be categorized into two groups. The first group is based on topological properties of non-overlapping community structure, while the second group uses the overlapping community structure features (i.e, a node could belong to multiple communities). We give a brief overview covering both types of strategies in order to highlight how communities can be advantageously used. However, experimental investigations are restricted to strategies designed for non-overlapping community structure which represents the focus of this study.

3.1 Local immunization strategies

These strategies target the most influential nodes using local information around randomly selected nodes. Their main advantage is that they require only a limited amount of information about the network topology. We present two local methods based on non overlapping community structure and one strategy designed for overlapping communities.

3.1.1 Community Bridge Finder (CBF)

Proposed by Salathe *et al.* [31], it is a random walk based algorithm designed to search for bridge nodes. The basic idea is that real-world networks exhibit a strong community structure with few links between the communities.

The CBF algorithm works as follows:

Step 1: Select a random node $v_{i=0}$ and follow a random path.

Step 2: $v_{i-1(i \geq 2)}$ is considered as a potential target if there is not more than one connection from v_i to any of the previous visited nodes.

Step 3: Two random neighboring nodes of v_i are picked (other than v_{i-1}). If there is no connections back to the previously visited nodes $v_{j < i}$ then, the potential target is marked as a bridge and it is immunized. Otherwise, a random walk at v_{i-1} is taken back.

Therefore, when a walker reaches a node in another community, he is no longer linked to previously visited sites. Comparisons have been performed with the Acquaintance strategy (A node is selected at random and one of its randomly selected neighbors is immunized). Extensive tests conducted on real-world and synthetic

249 networks using the SIR epidemic model show that CBF performs mostly better, of-
 250 ten equally well, and rarely worse than the Acquaintance strategy [13]. It performs
 251 particularly well on networks with strong community structure.

252 3.1.2 Bridge-Hub Detector (BHD)

253 The Bridge-Hub Detector [32] is another variant of CBF strategy. It targets bridge
 254 hub nodes for immunization by exploring friendship circles of visited nodes. The
 255 procedure of the BHD algorithm can be specified as follows:

256 Step 1: Select a random node $v_{i=0}$ and follow a random path.

257 Step 2: Let v_i be the node selected after i walks, and f_i be the set of all
 258 neighbors of the node v_i . The node v_i is targeted for immunization if there is at
 259 least a node in f_i that is not a member in the set F_{i-1} and that is not connected
 260 to the nodes in F_{i-1} where $F_{i-1} = f_0 \cup f_1 \cup f_2 \cup \dots \cup f_{i-1}$. Otherwise, v_i will not
 261 be targeted for immunization and F_i will be updated to $F_i = F_{i-1} \cup f_i$.

262 Step 3: One node v_H is randomly selected for immunization among the nodes
 263 in f_i that do not belong and could not be linked back to F_{i-1} .

264 Therefore, a pair of nodes, a bridge node and a bridge hub, are targeted for im-
 265 munization via a random walk. BHD was applied on simulated and empirical
 266 data constructed from social network of five US universities. Experimental results
 267 demonstrate that it compares favorably with Acquaintance and CBF strategies.
 268 Indeed, it results in reduced epidemic size, lower peak prevalence and fewer nodes
 269 need to be visited before finding the target nodes.

270 3.1.3 Random-Walk Overlap Selection (RWOS)

271 This random walk based strategy [33] targets the high degree overlapping nodes.
 272 The RWOS algorithm works as follows:

273 Step 1: Define the list of overlapping nodes L_{over} obtained from known or
 274 extracted communities.

275 Step 2: A random walk is followed starting from a random node $v_{i=0}$ of the
 276 network.

277 Step 3: The visited node v_i is nominated as a target for immunization if it be-
 278 longs to the list of overlapping nodes L_{over} , otherwise, the random-walk proceeds.
 279 Simulation results on synthetic and real-world networks with the SIR epidemic
 280 model show that the proposed method outperforms CBF and BHD strategies.
 281 In some cases it has a smaller epidemic size compared to the membership strat-
 282 egy where overlapping nodes are ranked according to the number of communities
 283 they belong to. In particular, its performance improves in networks with strong
 284 community structures and with greater overlap membership values.

285 3.1.4 Summary

286 Results show that local methods designed for networks with community structure
 287 are more efficient than classical local strategies. Key contributions of these works
 288 is to demonstrate that it is important to better take into account the modular
 289 organization of real-world networks in order to develop efficient immunization
 290 strategies. Note, however, that local methods are not as efficient as global ones.

291 Their main advantage is that they do not require a full knowledge of the global
292 structure of the network.

293 3.2 Global immunization strategies

294 Nodes are immunized according to a rank computed using a specific influence
295 (centrality) measure. Immunization aims to target nodes with high centrality due
296 to their big influence. The majority of known methods make use of the structural
297 information either at the microscopic or at the macroscopic level to characterize the
298 node importance. These strategies such as Degree and Betweenness immunization
299 strategies are very effective but they require the knowledge of the topology of the
300 entire network. Refer to [34] for a comprehensive survey on the subject. Given that
301 the influence of a node depends only on (i) the network's topology, and (ii) the
302 disease model, and that a vast majority of real-world networks exhibit a modular
303 organization, some global methods have been developed lately for such networks.

304 3.2.1 *Comm* strategy

305 Gupta *et al.* [35] proposed a new method called the *Comm* strategy. Nodes are
306 ranked using both the number of intra- and inter-community links, which respec-
307 tively link to nodes inside and outside the community. The purpose of this is
308 to rank nodes that are both hubs in their community and bridges between com-
309 munities. In this measure, the number of inter-community links is raised up to
310 power two while the number of the intra-community links is not raised to give
311 more importance to bridges. Results on synthetic and real-world networks show
312 that the *Comm* based strategy can be more effective than degree and betweenness
313 strategies. However, it gives significant importance to the bridges compared to the
314 community hubs. Yet, the hubs are commonly believed to be also influential nodes
315 as they can infect their many neighbors [36,37]. In some cases, they may play a
316 very major role in the epidemic spreading.

317 3.2.2 *Membership* strategy

318 Hebert-Dufresne *et al.* [38] proposed an immunization strategy based on the over-
319 lapping community structure of networks. Nodes are targeted according to their
320 membership number, which indicates the number of communities to which they
321 belong. Experiments with real-world networks of diverse nature (social, technolog-
322 ical, communication networks, etc.) and two epidemiological models show that this
323 strategy is more efficient as compared to degree, coreness and betweenness strate-
324 gies. Furthermore, its best performances are obtained for high infection rates and
325 dense communities.

326 3.2.3 *OverlapNeighborhood* strategy (*ON*)

327 Kumar *et al.* [39] proposed a strategy based on overlapping nodes. It targets im-
328 mediate neighbors of overlapping nodes for immunization. This strategy is based
329 on the idea that high degree nodes are neighbors of overlapping nodes. Using a

330 limited amount of information at the community structure level (the overlapping
 331 nodes), this strategy allows to immunize high degree nodes in their respective
 332 communities. Experiments conducted on four real-world networks show that this
 333 immunization method is more efficient than local methods such as CBF [40,41],
 334 BHD and RWOS methods. It also performs almost as well as degree and between-
 335 ness strategies while using less information about the overall network structure.

336 3.2.4 Community-Based Betweenness strategy (CBB)

337 In [42] Kitromilidis et al. define a strategy based on Community-based Between-
 338 ness measure, which is a redefinition of the standard Betweenness centrality. In
 339 this measure only paths that start and finish in different communities are taken
 340 into consideration. This strategy was used in order to characterize the influence of
 341 Western artists. It is based on the idea that an influential painter is the one who
 342 promotes the flow of ideas through different communities. Using a painter collab-
 343 oration network where links represent biographical connections between artists,
 344 they compared Betweenness with its classical version. Results show that the cbb
 345 performs better than the standard Betweenness. The modified centrality measure
 346 allows to highlight influential nodes who might have been missed as they do not
 347 necessary rank high in the standard measure.

348 3.2.5 Community-Based Mediator Strategy (CBM)

349 This immunization method is based on Community-Based Mediator measure [43].
 350 The idea behinds this strategy is that if an individual has many links in several
 351 communities, he can then play significant role to diffuse information around his
 352 circle. This method selects the most intermediate nodes which receive and dis-
 353 seminate information through the communities than other nodes. It combines the
 354 influence of the Degree and the Betweenness of the nodes in the network. The
 355 CBM measure is defined as follows:

$$CbM_i = H_i \frac{d_i}{\sum_{i=1}^N d_i} \quad (1)$$

356 Where H_i is the entropy of the node i . It is defined by the following formula:

$$H_i = \left[- \sum \rho_i^{in} \log(\rho_i^{in}) \right] + \left[- \sum \rho_i^{ex} \log(\rho_i^{ex}) \right] \quad (2)$$

357 Where ρ_i^{in} represents the fraction of links connected to i inside its community,
 358 while ρ_i^{ex} indicates the fraction of outgoing links from node i to nodes belonging
 359 to other communities. The entropy is used to find nodes that have a balance
 360 between the ability of diffusing the information in the network. The experimental
 361 results have shown that nodes with high CbM value have a greater impact to
 362 spread information in the network than nodes having a high Degree, Betweenness,
 363 CbC, PageRank or Eigenvector value.

3.2.6 Summary

Globally, experimental results demonstrate that the global strategies described above can reach the efficiency of classical strategies that are agnostic about the community structure while using less information. However, they do not distinguish between the various community structure strength that can be encountered in real-world networks (well defined, medium, loose). The Comm and Community-based Betweenness Strategies give more importance to nodes with a big amount of external links (the bridges). These nodes have a significant global influence in the network. Thus, these strategies are very efficient in networks with a community structure of medium strength. Indeed, the epidemics can propagate between communities through the high number of bridges in the network. Yet, these strategies are less efficient in networks with well-defined or loose community structure. Indeed, in these cases, the hubs can play a major role in the epidemic spreading process. The CBM strategy that immunizes nodes with a balance of external and internal links is more efficient in these situations. In order to overcome these drawbacks, we introduce three immunization strategies for networks with community structure strength ranging from well defined to loose community structure strength. Each strategy is tailored to one of the community structure strength of the network (well-defined, medium, loose). Moreover, they use also more information about the topological properties of the communities (the number of communities, community size and the density of inter-community links) to increase the performance of the community-based immunization strategies.

4 Proposed measures

In order to quantify the influence of a node in the diffusion process on community structured networks, we propose three measures that integrate various levels of information.

Let's $G(V, E)$ be a simple undirected network. V represents the set of nodes, and E is the set of edges. $C = \{C_1, \dots, C_k, \dots, C_m\}$ is the set of the non-overlapping communities while m is the number of communities of the network ($G = \bigcup_{k=1}^m C_k$).

4.1 Number of Neighboring Communities Measure

The main idea of this measure is to rank nodes according to the number of communities they reach directly (through one link). The reason for targeting these nodes is that they are more likely to contribute to the epidemic outbreak towards multiple communities. Note that all the nodes that do not have inter-community links share the same null value for this measure.

For a given node i belonging to a community $C_k \subset C$, the Number of Neighboring Communities $\beta_{NNC}(i)$ is given by:

$$\beta_{NNC}(i) = \sum_{C_l \subset C \setminus \{C_k\}} \bigvee_{j \in C_l} a_{ij} \quad (3)$$

Where a_{ij} is equal to 1 when a link between nodes i and j exists, and zero otherwise.

404 \bigvee represents the logical operator of disjunction, i.e., $\bigvee_{j \in C_l} a_{ij}$ is equal to 1 when
 405 the node i is connected to at least one of the nodes $j \in C_l$.

406 Some bridge nodes may be connected to a single neighboring community with
 407 a high number of inter-community links. Other bridge nodes may have a fewer
 408 amount of inter-community links but these links allow to reach multiple commu-
 409 nities. This strategy allows targeting nodes linked with a high number of external
 410 communities. Thus, it can target the most influential bridges of the network. In-
 411 deed, these nodes can disseminate epidemics to many communities all over the
 412 network. However, this strategy has some drawbacks. When the network has few
 413 communities, many nodes have the same rank. The nodes are randomly immunized
 414 in this case. Additionally, it does not target hubs for immunization. These nodes
 415 can affect a large number of nodes in their communities. They have then a major
 416 influence in their local communities. Overall, this strategy is suitable for networks
 417 with medium community structure having a high number of bridge nodes. Yet,
 418 it is inappropriate in networks with well-defined or loose community structure,
 419 where hubs may play a bigger role in the epidemic spreading.

420 4.2 Community Hub-Bridge Measure

421 Each node of the network share its links with nodes inside its community (intra-
 422 community links) and nodes outside its community (inter-community links). De-
 423 pending of the distribution of these links, it can propagate the epidemic more or
 424 less in its community or to its neighboring communities. Therefore, it can be con-
 425 sidered as a hub in its community and a bridge with its neighboring communities.
 426 That is the reason why we call this measure the Community Hub-Bridge measure.
 427 Furthermore, the hub influence depends on the size of the community, while the
 428 bridge influence depends on the number of its neighboring communities.

429 For a given node i belonging to a community $C_k \subset C$, the Community Hub-
 430 Bridge measure $\beta_{HB}(i)$ is given by:

$$\beta_{HB}(i)_{i \in C_k} = h_i(C_k) + b_i(C_k) \quad (4)$$

431 Where:

$$h_i(C_k) = \text{Card}(C_k) * k_i^{\text{intra}}(C_k) \quad (5)$$

432

$$b_i(C_k) = \beta_{NNC}(i) * k_i^{\text{inter}}(C_k) \quad (6)$$

433 $k_i^{\text{intra}}(C_k)$ and $k_i^{\text{inter}}(C_k)$ are respectively the intra-community degree and the
 434 inter-community degree of node i . $\text{Card}(C_k)$ is the size of its community. $\beta_{NNC}(i)$
 435 represents the number of its neighboring communities.

436 $h_i(C_k)$ tend to immunize preferentially hubs inside large communities. Indeed,
 437 they can infect more nodes than those belonging to small communities.

438 $b_i(C_k)$ allows to target nodes that have more links with various communities. Such
 439 nodes have a big inter-community influence.

440 The community Hub-Bridge strategy targets nodes that have a good balance
 441 between the intra-community and the inter-community links. It selects nodes play-
 442 ing simultaneously the role of hubs in their communities and bridges to other com-
 443 munities. This strategy gives the priority to hubs located in large communities due
 444 to their high local influence. These nodes can infect a big number of nodes in the

445 network if they are contaminated. Additionally, it targets bridges with the highest
 446 connectivity linked to the maximal number of external communities. This allows
 447 targeting nodes with the highest global influence in the network. However, this
 448 method gives importance to hubs as well as bridge nodes regardless of the commu-
 449 nity structure strength of the network. In some situations, more weight should be
 450 given to one of the two. For instance, in networks with non-cohesive communities
 451 the network act as one big community, in this case, more importance must be
 452 given to the hubs as they can infect several nodes in the network.

453 4.3 Weighted Community Hub-Bridge Measure

454 The Community Hub-Bridge measure targets in priority the hubs in large com-
 455 munities and the bridges linked to multiple communities. However, no importance
 456 is given to the community structure strength. When the community structure
 457 is well-defined, more importance should be given to the bridges. Indeed, in this
 458 case breaking the network in multiple communities allows to contain the epidemic
 459 spreading where it started. On the contrary, when the community structure is very
 460 loose, it is of prime interest to immunize the hubs in large communities. Weighting
 461 each component of the community Hub-Bridge allows therefore to give more or
 462 less importance to bridges or hubs according to the community structure strength.

463 For a given node i belonging to a community $C_k \subset C$, the Weighted Commu-
 464 nity Hub-Bridge Measure $\beta_{WHB}(i)$ is given by:

$$\beta_{WHB}(i)_{i \in C_k} = \rho_{C_k} * h_i(C_k) + (1 - \rho_{C_k}) * b_i(C_k) \quad (7)$$

465 Where ρ_{C_k} represents the interconnection density between the community C_k and
 466 the other communities of the network. It is given by:

$$\rho_{C_k} = \frac{\sum_{i \in C_k} k_i^{inter} / (k_i^{inter} + k_i^{intra})}{Card(C_k)} \quad (8)$$

468 If the communities are very cohesive, then more importance is given to the bridges
 469 in order to isolate the communities. Otherwise, more importance is given to the
 470 hubs inside large communities.

471 The epidemic diffusion of a node is dependent on its position in its community
 472 besides the relation that its community has with the other communities in the
 473 network. In this perspective, the Weighted Community Hub-Bridge is designed
 474 to be able to adapt with nodes belonging to communities with various structure
 475 strength. It is very similar to the Community Hub-Bridge strategy. Yet, it gives
 476 more weight to the bridges when the network has a well-defined community struc-
 477 ture for their isolation. Lets consider that an epidemic starts from the core of a
 478 community. If the community is isolated, then the epidemic stay confined in it and
 479 does not move to other parts of the network. This strategy gives also more weight
 480 to hubs in the case of networks with a very loose community structure since the
 481 network acts in this case as a single big community.

482 4.4 Toy example:

483 In order to illustrate the behavior of this measure a toy example is given in Fig-
 484 ure 2. Nodes are ranked according to the Number of Neighboring Communities
 485 measure in Figure 2 (a). Let's take the example of nodes $n5$ and $n10$ which are
 486 both community bridges and which have the same number of either internal and
 487 external links. According to Degree centrality measure in Figure 2 (d), both nodes
 488 have the same rank since it depends only on their number of neighbors. However,
 489 they have different ranks according to the Number of Neighboring Communities
 490 measure. The proposed measure gives more importance to node $n5$ which is linked
 491 to three external communities, so even if it is contaminated, it can transmit the
 492 epidemic disease first to its own community $C1$ and also towards the neighboring
 493 communities $C2$, $C3$ and $C4$. While the epidemic disease could be transmitted to
 494 nodes belonging to the communities $C1$ and $C2$ in the case of node $n10$ contamina-
 495 tion. Moreover, the nodes $n15$ and $n12$ are ranked among the less influential nodes
 496 according to Betweenness measure as it is shown in Figure 2 (e), although, both
 497 are community bridges that are likely to contribute to the epidemic outbreak to ex-
 498 ternal communities. Therefore, the Number of Neighboring Communities measure
 499 targets the most influential bridges which can spread the epidemics to multiple
 500 communities.

501 Figure 2 (b) shows the rank of nodes according to the Community Hub-Bridge
 502 measure. Even-though, both $n6$ and $n16$ have four inner links inside their own
 503 communities $n6$ is considered more influential because it is located in community
 504 $C1$ which is the largest community of the network. Therefore, it could be a threat
 505 to several nodes inside the network if ever it is infected. Unlike degree measure in
 506 Figure 2 (d) that classifies the nodes $n6$ and $n16$ in the same rank based on their
 507 number of connections without considering their location within the network. It is
 508 also noticed from Figure 2 (a) that many nodes have the same rank because they
 509 have the same number of neighboring communities. So, if we consider the nodes $n10$
 510 and $n12$, they are both connected to only one neighboring community (respectively
 511 $C1$ and $C3$), consequently they have the same rank. However, $n10$ has a bigger
 512 connectivity to $C1$ in term of the number of outer links. The reason why we
 513 introduced the quantity of outer links as a new parameter in the second term of
 514 the Community Hub-Bridge measure. This is to distinguish between bridges having
 515 big connectivity and those having lower connectivity with external communities.
 516 Based on Community Hub-Bridge measure $n10$ is more influential than node $n12$
 517 as it can be seen in Figure 2 (b) since it has three outer connections towards
 518 community $C1$ while node $n12$ has only one connection towards $C3$. Therefore, the
 519 influence of nodes according to this measure is linked to two factors: the importance
 520 of nodes inside their communities by giving the priority to those located in large
 521 communities, and the connectivity of the nodes towards various communities.

522 Nodes are ranked according to the Weighted Community Hub-Bridge measure
 523 in Figure 2 (c). The network given in this example has a well-defined community
 524 structure. As we can clearly see, if we take the example of the community $C1$,
 525 the density of inter-community links is equal to $\rho_{C_1} \approx 0.15$. Consequently, 15%
 526 of importance is given to the hub term $h_i(C_1)$ and 85% of importance is given
 527 to the bridge term $b_i(C_1)$. This explains why all the community bridges ($n5$, $n2$
 528 and $n4$) are immunized before the other nodes of the community $C1$. It helps to
 529 isolate this community and prevent the epidemic diffusion to move from $C1$ to the

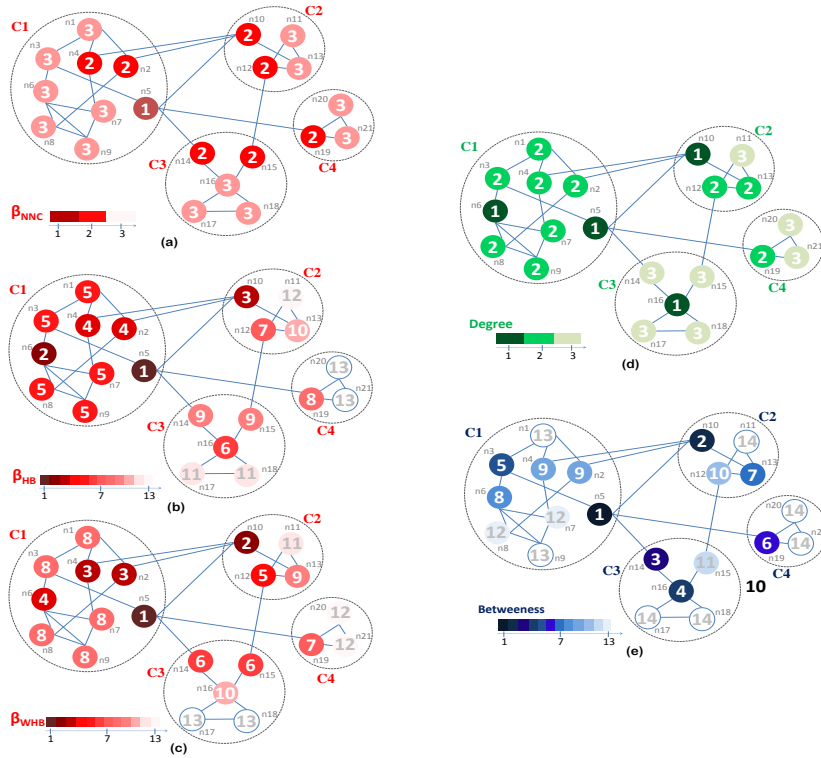


Fig. 2 Rank of nodes according to (a) Number of Neighboring Communities measure β_{NNC} (b) Community Hub-Bridge measure β_{HB} (c) Weighted Community Hub-Bridge measure β_{WHB} (d) Degree measure (e) Betweenness measure. Nodes are ranked from the most influential (nodes having the highest measure value) to the less influential node (nodes having the lowest measure value) in the network.

530 other communities of the network. Thus, the Weighted Community Hub-Bridge
 531 measure has the ability to adapt to the strength of the community structure. It
 532 gives more weight to the bridges when the network has a well-defined community
 533 structure in order to isolate the communities, while it gives more weight to hubs
 534 in the case of networks with a weak community structure since the network acts
 535 in this case like a single big community.

536 5 Experimental Setting

537 In this section, we present the data and methods used in the empirical evaluation
 538 of the various immunization strategies presented above.

539 5.1 Datasets

540 In order to evaluate the various measures under study, synthetic networks with controlled
 541 topological properties, together with real-world networks have been used.

Table 1 LFR network parameters

Number of nodes N	15 000
Average degree $\langle k \rangle$	7
Maximum degree k_{max}	122
Exponent for the degree distribution α	3
Exponent for the community size distribution σ	2.5
Mixing parameter μ	0.1, 0.4, 0.7, 0.9
Community size range set	[50 250],[100 500]

5.1.1 Synthetic networks

Synthetic networks are generated using the LFR (Lancichinetti, Fortunato and Radicchi) algorithm [44]. It generates random samples of networks with power-law distributed degree and community size. Hence, LFR algorithm guarantees networks with realistic features [45]. This algorithm allows to control different parameters when generating networks. Mainly, the mixing parameter μ , determines the ratio of the number of external neighbors of a node to the total degree of the node. Its value controls the strength of the community structure. For small values of μ , the communities are well-separated because they share few links, whereas when μ increases the proportion of inter community links becomes higher, making community identification a difficult task. Experimental studies showed that for a scale-free network, the degree distribution exponent α usually ranges from 2 to 3, and the maximal degree is estimated to be $k_{max} \sim n^{1/(\alpha-1)}$ [46,47,48]. The parameters values used in our experiments are given in Table 1.

5.1.2 Real-world networks

Real-world networks of various nature (online social networks, a technological network and a collaboration network) are used in order to test the immunization strategies.

- **Facebook:** We use a network gathered by Traud *et al.* [49] from Facebook¹ online social network. This data includes the friendship network of five universities in the US. It provides also information about the individuals such as the dormitory, the major or the field of specialization and the year of class.

- **Power-grid:** This technological network is an undirected, unweighted network containing information about the topology of the Western States Power Grid of the United States. An edge represents a power supply line. A node is either a generator, a transformer or a substation. This data² is compiled by D. Watts and S. Strogatz [50].

- **General Relativity and Quantum Cosmology (GR-QC):** GR-QC³ is a collaboration network collected from the e-print arXiv. It covers scientific collaborations between authors of papers submitted to the General Relativity and Quantum Cosmology category. The nodes represent the authors and there is a link between two nodes if they co-authored a paper. This data is available in the SNAP repository compiled by Leskovec *et al.* [51].

¹<http://code.google.com/p/socialnetworksimulation/>

²<http://www-personal.umich.edu/~mejn/netdata/>

³<http://snap.stanford.edu/data/ca-GrQc.html>

Table 2 The basic topological properties of six real-world networks. N and E are respectively the total numbers of nodes and links. Q is the modularity. N_c is the number of communities. λ_{th} is the epidemic threshold.

Network	N	E	Q	N_c	λ_{th}
Caltech	620	7255	0.788	13	0.012
Princeton	5112	28684	0.753	21	0.006
Georgetown	7423	162982	0.521	42	0.006
Oklahoma	10386	88266	0.914	67	0.031
Power grid	4941	6594	0.92	41	0.092
CR-QC	5242	14496	0.86	396	0.059

575 As the community structure of these networks is unknown, we use a community
 576 detection algorithm. We choose to use the Louvain algorithm that proved to be
 577 efficient in both synthetic and real-world networks [23,52]. Furthermore, the topo-
 578 logical properties of the uncovered communities are also realistic [53].

579 The basic topological properties of these networks are given in Table 2.

580 5.2 SIR model parameters

581 The value of the transmission rate λ is chosen to be greater than the network
 582 epidemic threshold λ_{th} in order to better characterize the spreading capability, it
 583 defined as [54]:

$$\lambda_{th} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \quad (9)$$

584 Where $\langle k \rangle$ and $\langle k^2 \rangle$ are respectively the first and second moments of the
 585 degree distribution. The epidemic threshold values λ_{th} of all the networks used in
 586 this paper are reported in Table 2. The same transmission rate value ($\lambda = 0.1$) is
 587 used in all the experiments. It is larger than the values of the epidemic threshold
 588 λ_{th} of all the data collection used in this work. We set also the value of the
 589 recovery rate γ to 0.2. This small value is chosen in order to give each infected
 590 node many chances to infect its neighbors with the probability γ before changing
 591 to the recovered status.

592 5.3 Immunization scheme

593 To investigate the spread of an infectious disease on a contact network, we use the
 594 methodology described in Figure 3. For Global strategies, the influence of every
 595 node in the network is calculated according to a given centrality measure. Then,
 596 nodes are sorted in decreasing order of their influence values. Next, nodes with
 597 highest centrality are removed from the network (or their state is set to resistant)
 598 until a desired immunization coverage is achieved. For Local immunization, nodes
 599 are targeted and removed according to a random strategy initiated from randomly
 600 chosen nodes in the network. In both cases, the network obtained after the tar-
 601 geted immunization is used to simulate the spreading process, running the SIR
 602 epidemic model simulations. After a simulation, we record the total number of
 603 cases recovered (the epidemic size). In order to ensure the effectiveness of the SIR

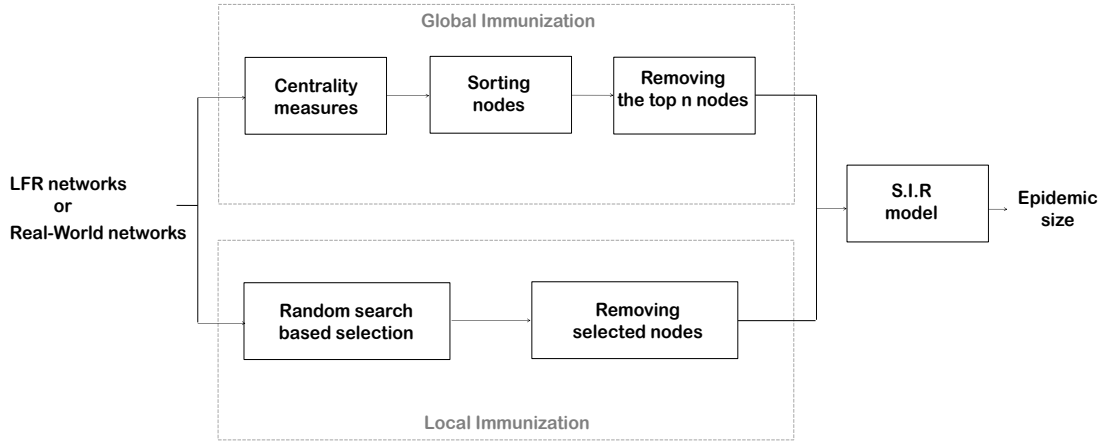


Fig. 3 The main steps of the immunization scheme.

604 propagation model evaluation, results are averaged over 600 independent realiza-
 605 tions. Finally, we calculate the mean epidemic size to evaluate the effectiveness of
 606 the proposed methods.

607 5.4 Evaluation Criteria

608 5.4.1 Relative difference of outbreak size

609 To compare the performance of different immunization strategies, we use the frac-
 610 tion of the epidemic size. We also use the relative difference of outbreak size $\Delta r_{\beta_0, \beta_c}$
 611 defined by:

$$\Delta r_{\beta_0, \beta_c} = \frac{R_{\beta_0} - R_{\beta_c}}{R_{\beta_0}} \quad (10)$$

612 Where R_{β_0} and R_{β_c} are respectively the final numbers of recovered nodes for the
 613 alternative and the proposed strategy after the SIR simulations. If the relative
 614 difference of outbreak size is positive, the epidemic spreads less with the proposed
 615 strategy. Therefore, it is the most efficient one. Otherwise, the epidemic spreads
 616 more with the proposed strategy and the alternative strategy is more efficient.

617 5.4.2 Largest Connected Component

618 We use also the size of the Largest Connected Component *LCC* to test the effec-
 619 tiveness of the proposed strategies regardless of the epidemiological models used. It
 620 is the largest remaining subgraph after the simulation process. The size of the
 621 largest connected component is used to measure the maximum limit to which an
 622 epidemic can spread. The *LCC* is one of measures to quantify the performance
 623 of ranking strategies. It focuses on the changes of the structure of the giant com-
 624 ponent after removing some nodes. In effect, the size of the largest connected

625 component is computed after removing a certain proportion of nodes selected ac-
626 cording to an immunization strategy. Clearly, the smaller the LCC , the better the
627 immunization strategy.

628 **6 Results and discussion**

629 In this section, we report the results of two sets of experiments. The first set
630 of experiments is performed with synthetic networks with controlled community
631 structure. It is aimed at getting a better understanding of the relationship between
632 the community structure and the centrality measures. These experiments are con-
633 ducted on networks generated with the LFR algorithm. Indeed, this algorithm
634 allows to control various topological properties of the community structure. We
635 investigate the influence of the strength of the community structure. The commu-
636 nity size range effect is also studied. Finally, the proposed immunization strategies
637 are compared with both global strategies (Degree, Betweenness and Comm strate-
638 gies) and local strategies (Community Bridge Finder [31] and Bridge Hub Detector
639 [32] strategies).

640 The second set of experiments concerns real-world networks. Online Social net-
641 works, a technological network and a collaboration network are used. Recall that,
642 as there is no ground-truth data for these networks, the community structure is
643 uncovered using the Louvain Algorithm. Indeed, previous studies on synthetic net-
644 works have shown that it succeeds in identifying the communities for a large range
645 of community structure strength [52]. First, the proposed immunization strategies
646 are compared and discussed, then their evaluation is performed against both local
647 and global alternative strategies.

648 **6.1 Synthetic networks**

649 *6.1.1 Influence of the strength of the Community Structure*

650 In the LFR model, the mixing parameter value μ varies from 0 to 1. It allows
651 to control the strength of the community structure from well-separated commu-
652 nities with few inter-community links (low values of μ) to a network with no
653 community structure (high values of μ). In order to investigate the effect of the
654 strength of the community structure on the performance of the proposed meth-
655 ods, five networks have been generated for each value of the mixing parameter
656 ($\mu = 0.1, 0.4, 0.7$ and 0.9). Figure 4 reports the average fraction of the epidemic
657 size versus the proportion of immunized nodes for each μ value. According to the
658 results reported in this figure, the performance of both Community Hub-Bridge
659 and Weighted Community Hub-Bridge strategies decreases while increasing the
660 mixing parameter value. Whatever the fraction of immunized nodes, the meth-
661 ods perform best when the communities are well-separated. When the fraction of
662 inter-connections between the communities increases, performance decrease grad-
663 ually. Indeed, with well-separated communities, the epidemics is localized to few
664 communities, while it tends to spread more when the inter connections increase.
665 The Number of Neighboring Communities strategy shows its best performance
666 for a medium range community strength value ($\mu = 0.4$). Its efficiency decreases

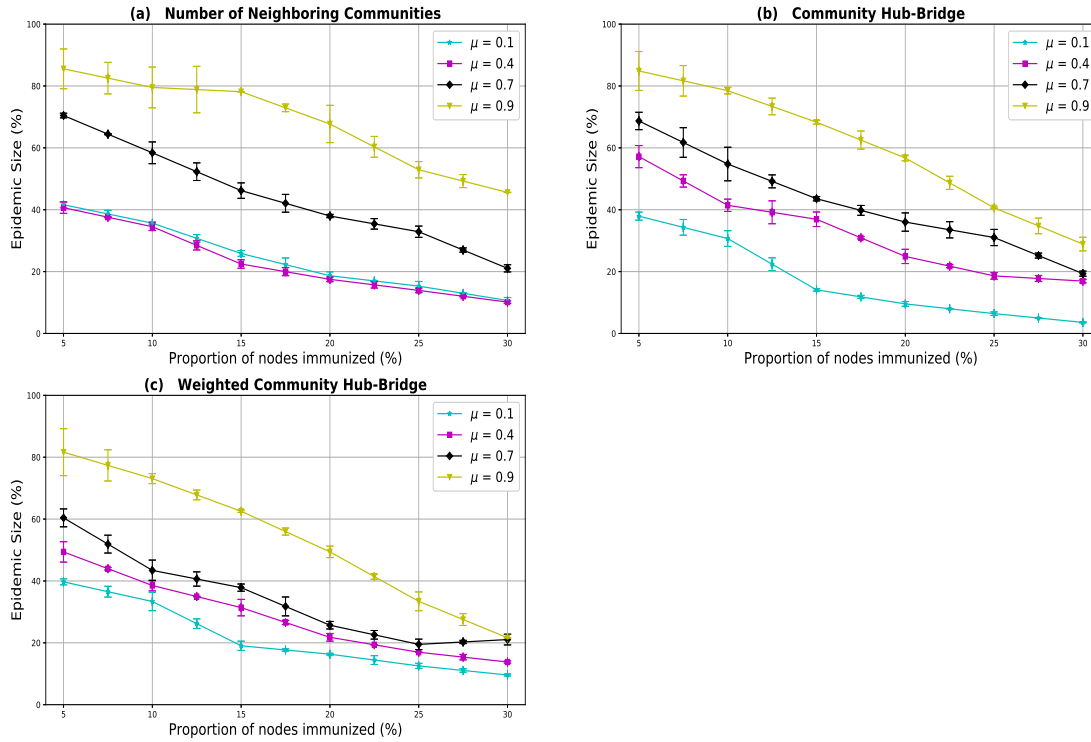


Fig. 4 Influence of the strength of the community structure on the epidemic size of the proposed methods. Each point show the epidemic size with respect to the fraction of the immunized nodes. Simulations are performed on LFR-generated networks with various mixing parameter values μ . Each epidemic size value is the average of 600 S.I.R simulation runs.

667 slightly in the case of well-defined community structure, and it gets even worse
 668 when it is very loose. Let's now turn to the comparisons of the proposed meth-
 669 ods between them. We can distinguish three cases depending of the community
 670 structure strength.

671 In networks exhibiting a very strong community structure, we can see in Fig-
 672 ure 4 that the Community Hub-Bridge strategy is the most efficient. This is due to
 673 the fact that both alternatives methods (Number of Neighboring Communities and
 674 Weighted Community Hub-Bridge strategies) target preferentially the bridges. In
 675 fact, it is not the best solution in a network where the intra-community links pre-
 676 dominate. As there is few external connections compared to the total connections
 677 (intra-community links are considered to be 90% of the total links of the network
 678 when $\mu=0.1$), local outbreaks may die out before reaching other communities.
 679 Therefore, immunizing community hubs seems to be more efficient than immuniz-
 680 ing bridges in networks with strong community structure. This is the reason why
 681 the Community Hub-Bridge method which targets nodes having a good balance
 682 of inner and outer connections is more efficient.

683 In networks with weak community structure as it can be seen in Figure 4, the
684 Weighted Community Hub-Bridge method is the most efficient. Indeed, when μ
685 has a high value, the network does not have a well-defined community structure.
686 In that case, Weighted Community Hub-Bridge strategy can better adapt to the
687 community structure. It gives more weight to the community hubs as they are the
688 most influential nodes in networks with a loose community structure. Remember
689 that the network acts as a single community in the extreme case where $\mu \geq 0.9$.
690 That is the reason why it performs better than the other proposed methods.

691 In networks with community structure of medium strength, the Number of
692 Neighboring Communities outperforms all the other proposed methods as it re-
693 ported in Figure 4. In this type of networks, nodes have many external connections
694 while maintaining a well-preserved community structure. Therefore, there is much
695 more options for the epidemic to spread easily to neighboring communities. As the
696 Number of Neighboring Communities strategy targets the most influential com-
697 munity bridges, it prevents the epidemic spreading to multiple communities. This
698 is the reason why this immunization method shows its best performance in this
699 case.

700 To summarize, Community-Hub bridge strategy is well-suited to situations
701 where the communities are well-defined (Dense communities with few links be-
702 tween communities). The Weighted Hub-Bridge strategy is recommended when
703 the community structure is very loose. For situations in between, the Number of
704 Neighboring Communities strategy is more efficient.

705 *6.1.2 Community Size Range effects*

706 The aim of this investigation is to show the impact of the community size range on
707 the performance of the proposed methods. Studies reported above have been per-
708 formed with community structure size in the range [100, 500]. In this paragraph,
709 they are also evaluated in networks with community size range equals to [50, 250].
710 Figure 5 reports the epidemic size versus the percentage of immunized nodes for
711 values of the mixing parameter μ ranging from $\mu = 0.1$ to $\mu = 0.9$, and with
712 the two community size range under study. One can see that all the immunization
713 strategies exhibit the same behavior. They always perform better in networks with
714 smaller community size range. Furthermore, the differences between the epidemic
715 sizes in the two situations decrease when the proportion of immunized nodes in-
716 creases. In networks with a big community size range, there are a small number
717 of communities. Consequently, the range of the Number of Neighboring Commu-
718 nities measure is also small, and many nodes have the same values (as it is shown
719 in the example given in section 3.1). That makes the ranking less efficient. In net-
720 works with a smaller community size range, there are much more communities. In
721 this case, more nodes have different numbers of neighboring communities values
722 and the ranking is more efficient. That is the reason why the Number of Neigh-
723 boring communities performs better in the latter case. Concerning Community
724 Hub-Bridge and Weighted Community Hub-Bridge measures, both are weighted
725 by the number of neighboring communities. This weight becomes more discrimina-
726 tive as the community size range decreases. That explains why they also perform
727 better in networks with small community size range. Additionally, it is illustrated
728 also in Figure 5 (d)-(f) that the other community based strategies (Comm, CBB
729 and CBM) are not affected by the community size range. To confirm that, we

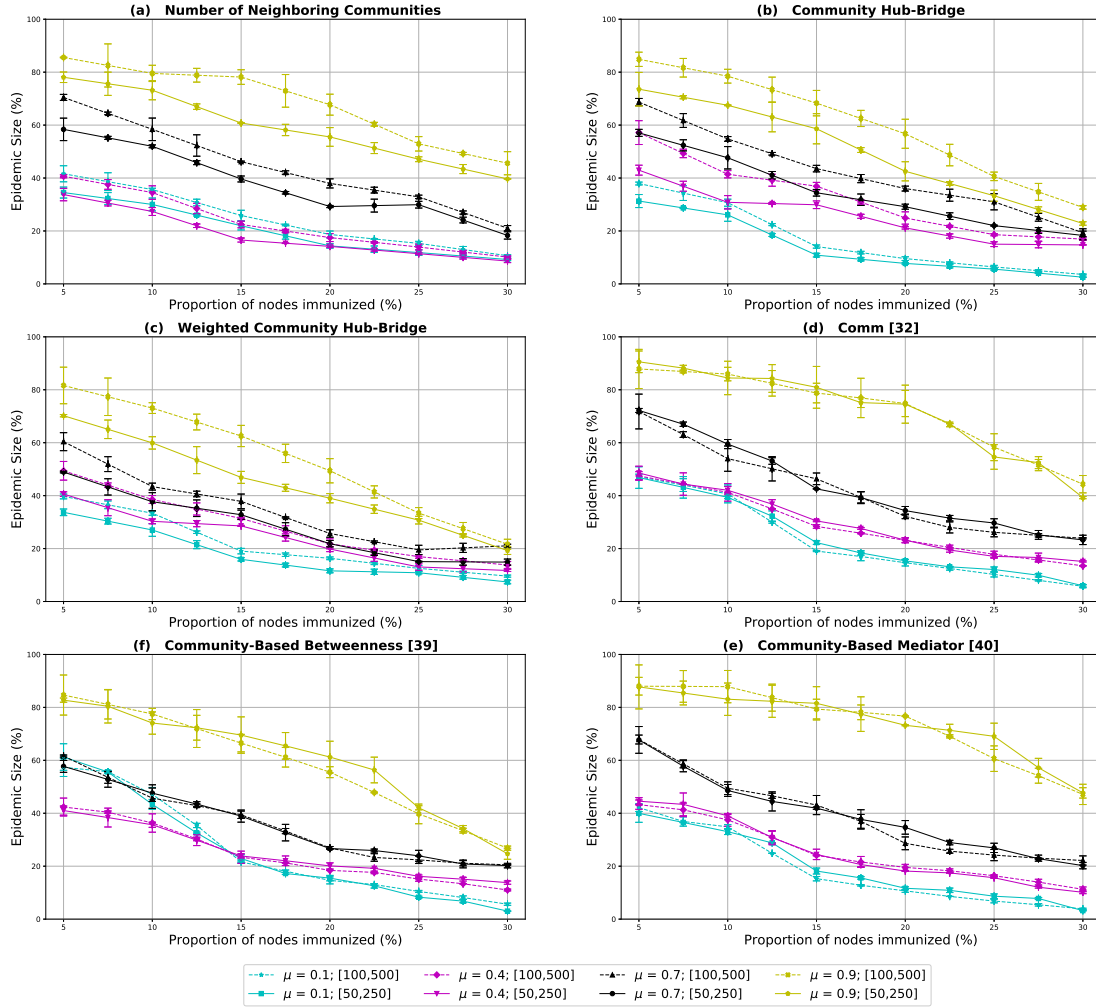


Fig. 5 Effect of community size range on the epidemic spreading of the proposed methods. Each point show the epidemic size as a function to the fraction of the immunized nodes. Simulations are performed on LFR network with different community structure. Each value is the average of 600 runs per network and immunization method.

730 perform the Analysis of Variance ANOVA on the performance of these strategies
 731 while varying both the community size range and the mixing parameter of the
 732 network. Based on the results reported in Table 6, one can notice that the esti-
 733 mated F value is always smaller than the critical value of F ($F < F_{critic}$). Thus,
 734 we can conclude that Comm, CBB and CBM exhibit the same performance while
 735 changing the community size range, and this in networks with various community
 736 structure strength. These methods are based on measures that do not take into

737 account either the size and the number of communities. Hence the stability of their
 738 performance despite the change of the community size range.

739 From Figure 5, one can also see that the Community Hub-Bridge method is
 740 always the best immunization method in networks with well-defined community
 741 structure, and that the Number of Neighboring Communities outperforms the
 742 other proposed immunization methods (where $\mu = 0.4$). Moreover, the Weighted
 743 Community Hub-Bridge is still the most efficient one in networks with non-cohesive
 744 community structure.

745 6.1.3 Comparison with the alternative methods

746 Figure 6 reports the relative difference of the outbreak size between the proposed
 747 strategies and both local (Acquaintance, CBF, BHD) and global alternatives (De-
 748 gree, Betweenness, Comm, Community-based Betweenness and Community-based
 749 Mediator) as a function of the fraction of immunized nodes. Community Hub-
 750 Bridge is taken as the reference in (a), (d) and (g), Number of Neighborhood
 751 communities in (b), (e) and (h) and Weighted Community Hub-Bridge in (c), (f)
 752 and (e). The values of the mixing parameter ($\mu = 0.1$, $\mu = 0.4$, $\mu = 0.7$) cover
 753 the three situations in terms of community strength (strong, median and weak
 754 community structure).

755 Figure 6 (a), (d) and (g) shows that $\Delta R/R_0$ has usually a positive value. Thus,
 756 Community Hub-Bridge yields a smaller epidemic size compared to all the alter-
 757 native methods whatever the fraction of immunized nodes values, and this holds
 758 for all the range of community structure strength. The middle panels of Figure 6
 759 reports the results of the comparative evaluation of the Number of Neighboring
 760 Communities strategy. Overall, it is more efficient than the tested alternative
 761 methods. However, Betweenness and Community-based Mediator perform better
 762 in networks with strong community structure ($\mu = 0.1$). Indeed, the relative differ-
 763 ence is negative in this case. Therefore, targeting the community bridges is not the
 764 best immunization solution in networks with very well-defined community struc-
 765 ture. It performs also worse than the Community-based Betweenness in networks
 766 with loose community structure ($\mu = 0.7$). It can be also noticed from Figure 6 (c),
 767 (f) and (i) that the Weighted Community Hub-Bridge method results always in the
 768 lowest epidemic size compared to the other methods. To summarize, if we exclude
 769 the case of the Number of neighborhood Communities strategy in the situation
 770 where the network has a strong community structure ($\mu = 0.1$), in every other
 771 situations the relative difference of the outbreak is always positive. That indicates
 772 that all the proposed strategies outperform the alternatives. Let's now turn to more
 773 detailed comparisons. First of all, these results clearly demonstrate the superiority
 774 of global methods over local methods. Indeed, in any case, the biggest differences
 775 are observed with Acquaintance followed by CBF and BHD. In fact, their rank is
 776 correlated with the level of information that they possess on the network topol-
 777 ogy. In fact, Acquaintance is totally agnostic about the network topology, CBF
 778 targets the bridges between the communities while BHD targets both bridges and
 779 hubs. Even though CBF and BHD are community-based methods, they use the
 780 information only at the level of randomly chosen nodes, from where their low
 781 performances. The compared effectiveness of the five alternatives global strategies
 782 depends on the strength of the community structure. For strong community struc-
 783 ture ($\mu = 0.1$), Degree and Comm strategy are very close while Betweenness and

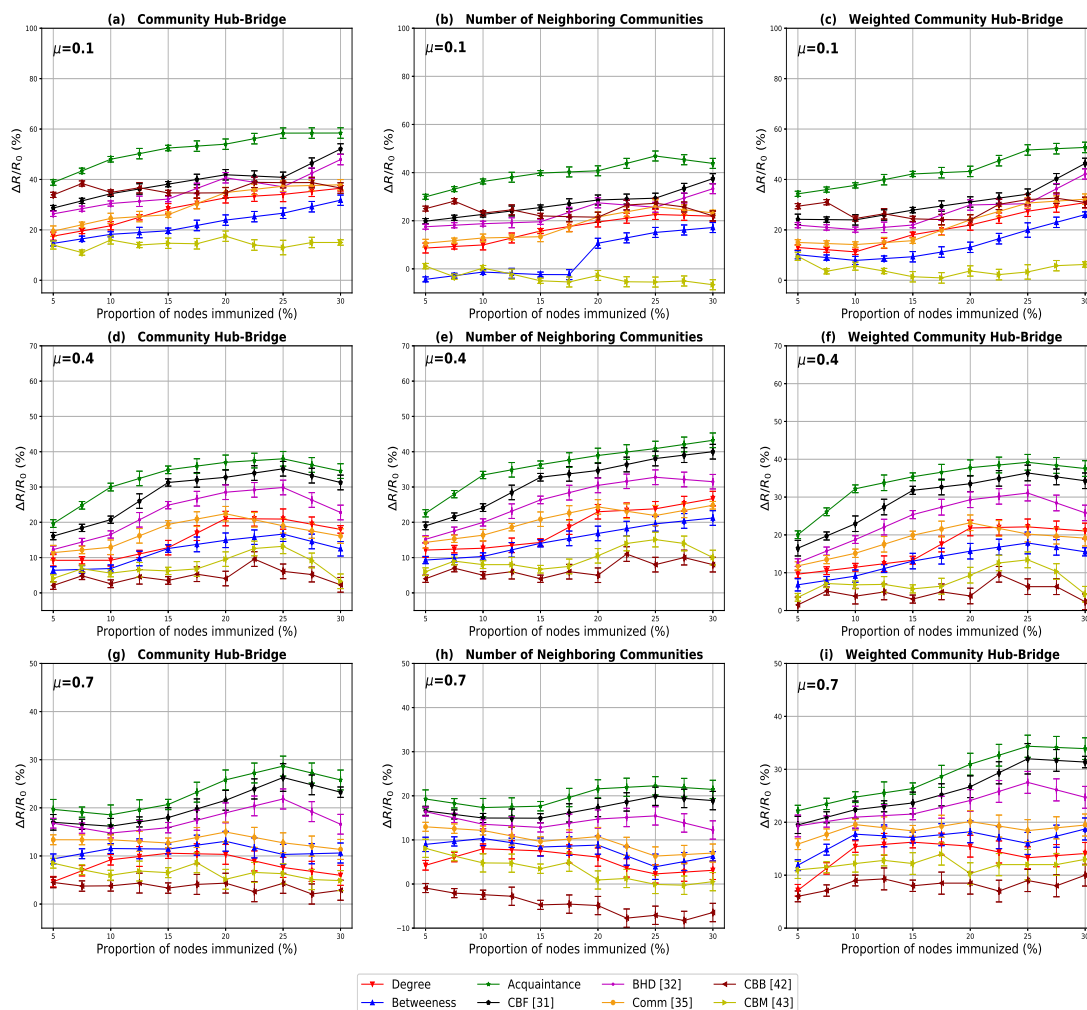


Fig. 6 The relative difference of the outbreak size $\Delta R/R_0$ as a function of the fraction of immunized nodes. The left panels show the difference between Community Hub-Bridge method and the alternative methods. The middle panels show the difference between Number of Neighboring Communities method and the alternative methods, while the right panels show the difference between Weighted Community Hub-Bridge and the alternative methods. We note that a positive value of $\Delta R/R_0$ means a higher performance of the proposed method. Simulations are performed on LFR network with different community structure. Final values are obtained by running 600 independent simulations per network, immunization coverage and immunization method.

784 Community-based Mediator methods are slightly more performing whatever the
 785 value of fraction of immunized nodes. For medium community structure strength
 786 ($\mu = 0.4$), results are more mixed, even if Community-based Betweenness is still a

Table 3 The estimated mixing parameter μ of the real-world networks.

Network	Power-grid	ca-GrQc	Princeton	Oklahoma	Caltech	Georgetown
μ	0.034	0.095	0.354	0.441	0.448	0.522

787 little bit more efficient. For weak community structure ($\mu = 0.7$), the five strate-
 788 gies are well separated. Community-based Betweenness ranks first, followed by the
 789 Community-based Mediator, Degree, Betweenness and Comm strategy in terms of
 790 efficiency.

791 6.2 Real-world networks

792 As our goal is to cover a wide range of situations, real-world data come from
 793 different domains (social, technical and collaboration networks). In order to link
 794 the results of this set of experiments with those performed on synthetic data, we
 795 estimate the mixing proportion parameter of the uncovered community structure
 796 by the Louvain algorithm. Indeed, experiments performed with synthetic networks
 797 have shown that the community strength is a major parameter in order to explain
 798 the efficiency of the proposed immunization strategies. Estimated values reported
 799 in Table 3 show that the networks cover a wide range of community strength.

800 6.2.1 Spreading efficiency of the proposed methods

801 Figure 7 shows the epidemic size as a function of the fraction of immunized nodes
 802 obtained after the SIR simulations for the proposed immunization methods. These
 803 results corroborate the conclusions we made with the synthetic networks. It can
 804 be observed on Figure 7 (e) and (f) that in networks with strong community struc-
 805 ture, Community Hub-Bridge is the most efficient immunization method. Indeed,
 806 the estimated mixing parameter value μ is equal to 0.03 and 0.09 respectively for
 807 the power-grid network and the collaboration network. Communities are very well
 808 separated, and the Community Hub-Bridge method targets nodes with a good bal-
 809 ance of intra-community and inter-community links. That is where its superiority
 810 lies.

811 In networks with average community structure strength shown in Figure 7
 812 (a), (b) and (d), the Number of Neighboring Communities outperforms the other
 813 proposed methods. It targets the bridges connected to multiple communities which
 814 facilitates the spread of epidemics throughout the whole network. Therefore, it is
 815 the most efficient method in Caltech, Princeton and Oklahoma networks.

816 The Weighted Community Hub-Bridge is the most efficient method for the
 817 Georgetown network (where $\mu = 0.522$) as reported in Figure 7 (c). This method
 818 depends on the fraction of the inter-community links for each community within
 819 the network, which allows us to give the appropriate weighting to favor either
 820 the inter-community or the intra-community influence. This is the reason why
 821 it outperforms the other proposed methods in the Georgetown network which
 822 does not have a strong community structure. Finally, these results confirm the
 823 paramount influence of the mixing proportion parameter in order to choose the
 824 most appropriate strategy in a given situation. Based on the above results obtained

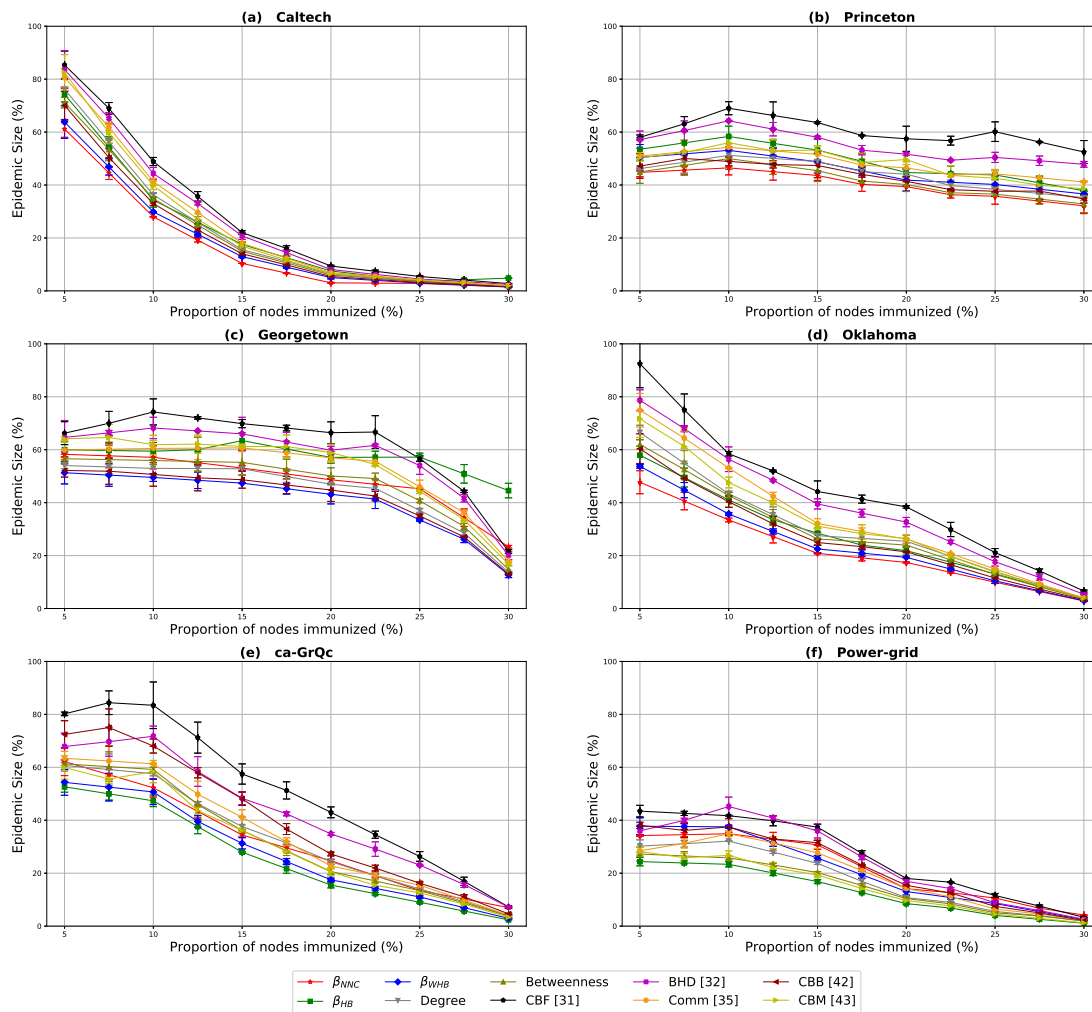


Fig. 7 The epidemic size of the immunization methods performed on six real networks of different types namely on facebook network of four universities (a) Caltech (b) Princeton (c) Georgetown (d) Oklahoma, and on (e) Collaboration network (f) Power network. Final values are obtained by running 600 independent simulations per network, immunization coverage and immunization method.

825 after using real-world networks with different structures, sizes and types, what
 826 matters the most is the strength of the community structure.

6.2.2 Comparison with the alternative methods

The relative difference of the outbreak size between global as well as local strategies and the proposed strategies is reported in Figure 8. Similarly, with synthetic networks, the local strategies (CBF and BHD) perform poorly as compared to global ones. Indeed, it appears clearly that these two types of methods are well separated. The results of the comparative evaluation of the global strategies are quite consistent with what might have been expected. The proposed strategies are globally more efficient than their competitors. This is all the more true when they are used appropriately.

The left panels of Figure 8 show the comparison between the Community Hub-Bridge and the alternative methods. It outperforms the global methods in networks with strong community structure (Power Grid and ca-GrQc) with a minimal gain of around 10 % over the best alternatives (Community-based Mediator and Betweenness). Its benefits reduces when the strength of the community structure gets looser. It is still above Betweenness for Princeton and Oklahoma networks, despite their medium range mixing proportion. However, when the community structure becomes weaker (Caltech and Georgetown), it is less performing than both Community-based Mediator and Betweenness when the fraction of immunized nodes is greater than 20 %.

The middle panels of Figure 8 show the comparison between the Number of Neighboring Communities and its alternative. It shows its best performances for networks with medium mixing proportion values (Princeton, Oklahoma, Caltech) with gains above 10 % as compared to the most performing alternative (Community-based Betweenness). However, it performs less than Community-based Betweenness while it is still performing better than the other alternatives for Georgetown (such as degree strategy with gains of less than 10 %). However, it performs in some cases worse than Community-based Mediator, Degree and Betweenness in networks with strong community structure (Power-grid and the collaboration networks).

The right panels of Figure 8 show the comparison between the Weighted Community Hub-Bridge and the alternative strategies. As expected, it outperforms its competitors in networks with average and high community structure strength. However, it can be worse than Community-based Mediator and Betweenness for networks with a strong community structure.

To summarize, these experiments reveal that the proposed algorithms are very effective in identifying the influential nodes to be selected for immunization. When they are used on the appropriate networks in terms of community strength, they outperform the available strategies, simply by using relevant information about the community structure.

6.2.3 Influence of the diffusion process parameters

To test the robustness of the results to the variation of the SIR parameters model, simulation results with $\lambda = 0.9$ and $\gamma = 0.2$ are reported in Figure 9. Results, show that increasing the infection rate λ , a greater proportion of immunized nodes is needed to mitigate the spread of the epidemic. This is valid for all the tested immunization methods. For instance, only 30% of the nodes need to be removed (immunized) for all the strategies to stop the epidemic spreading in the power-grid

Table 4 The Normalized Mutual Information *NMI* in Power-grid and Georgetown networks.

Network	Power-grid			Georgetown		
NMI	Louvain	WalkTrap	Infomap	Louvain	WalkTrap	Infomap
Louvain	-	0.872	0.751	-	0.287	0.181
WalkTrap	-	-	0.818	-	-	0.429
Infomap	-	-	-	-	-	-

873 network when λ is equal to 0.1 (Figure 7 (f)), while around 50% of the nodes need
874 to be immunized when λ is equal to 0.9 (Figure 9 (a)). In the Georgetown network,
875 around 40% of nodes need to be immunized when λ is equal to 0.1 (Figure 7 (c)),
876 while a 60% node immunization rate is required in the case of a high infection rate
877 value(Figure 9 (b)), and this hold for all the strategies. Therefore, the probability
878 that an infected node contaminates its neighbors gets higher with an increase of
879 the infection rate λ . Thus, the epidemic spreads at a higher rate. Consequently,
880 a bigger proportion of immunized nodes is needed to prevent the spread of the
881 epidemic.

882
883 We also employ the size of the Largest Connected Component *LCC* to confirm
884 the effectiveness of the proposed strategies. Figure 10 reports the *LCC* of vari-
885 ous immunization methods computed on two real-world networks with different
886 community structure strength (Power-grid and Georgetown network). For both
887 networks in Figure 10, one can see that increasing the proportion of immunized
888 nodes, the size of the largest connected component declines. In the power-grid
889 network, the curve of the Community Hub-Bridge strategy declines faster than all
890 the other alternative global and local strategies, as it is reported in Figure 10 (a).
891 Thus, the network can be broken down efficiently by selecting the influential nodes
892 according to this strategy. It is followed by the Weighted Community Hub-Bridge,
893 and the Community-based Mediator strategy. The local methods (Community
894 Bridge Finder and Bridge-Hub Detector) perform poorly to split the network.
895 The Weighted Community Hub-Bridge strategy is the most effective strategy in
896 Georgetown network as it is reported in Figure 10 (b). Its curve declines faster
897 than all the other immunization strategies before reaching the steady state. There-
898 fore, immunizing nodes according to the Weighted Community Hub-Bridge makes
899 the network split quickly into several independent modules, which leads to its col-
900 lapse. As in the case of the SIR model, the Community Hub-bridge strategy is the
901 most effective strategy in networks with a well-defined community structure (e.g.,
902 power-grid network), while the Weighted Community Hub-Bridge strategy outper-
903 forms all the other methods in networks exhibiting a loose community structure
904 (e.g., Georgetown network).

905 6.3 Influence of the community detection algorithm

906 In this section, we report a set of experiments on Power-grid and Georgetown net-
907 works using WalkTrap and Infomap community detection algorithms. We choose
908 these two networks because Power-Grid has a well-defined community structure
909 while Georgetown has a loose community structure. The aim of these experiments
910 is to get a clear picture about the community structure variations.

Table 5 The estimated mixing parameter μ , the number of communities N_c and the modularity Q in Power-grid and Georgetown networks.

Network	Metric	Detection algorithm		
		Louvain	WalkTrap	Infomap
Power-grid	μ	0.034	0.036	0.039
	N_c	41	45	53
	Q	0.92	0.907	0.931
Georgetown	μ	0.522	0.515	0.487
	N_c	42	193	272
	Q	0.521	0.546	0.604

911 6.3.1 Community detection algorithms comparison

912 To compare the community structure uncovered by WalkTrap, Infomap and Lou-
913 vain, we use the Normalized Mutual Information NMI as it is commonly used in
914 the community detection literature [55]. Their estimated values for each network
915 are reported in Table 4. In the network with a well-defined community struc-
916 ture (e.g., Power-grid network), the NMI values are high. This means that the
917 community structures uncovered by the three algorithms are very similar. In the
918 Georgetown network NMI values are below 0.5. This indicates that the community
919 structures are quite different.

920 We also report the proportion of inter-community links, the number of de-
921 tected communities, and the modularity value in Table 5. For the network with
922 a well-defined community structure, the three algorithms detect nearly the same
923 number of communities with a relatively larger number for Infomap . This confirm
924 the similarity of the community structure. For the Georgetown network we ob-
925 serve a large variation of this parameter. This is another sign that the community
926 structures are very dissimilar. The Modularity measures the quality of the commu-
927 nity structure. Its values are very high when the network community structure is
928 well-defined, and relatively low for the network with a loose community structure.
929 According to this parameter, the performance of the three algorithms are compa-
930 rable for networks with well-defined communities. Infomap is the most accurate
931 algorithm, followed by WalkTrap then Louvain when the community structure is
932 loose.

933 To summarize, when the community structures are well-defined (low values of
934 the proportion of the inter-community links) the algorithms uncover about the
935 same communities, while when the community structure is loose their results can
936 be quite different. Furthermore, the mixing parameter values using the different
937 algorithms are very close. Globally, the three detection algorithms have the same
938 performance in networks with well defined communities with a slight preference
939 for Infomap.

940 6.3.2 Influence of the community detection algorithms on the proposed methods

941 Figure 11 represents the epidemic size of the proposed strategies versus the pro-
942 portion of immunized nodes for different community detection algorithms. The
943 immunization methods are tested on two networks with different community struc-
944 ture strength. This figure shows the effect of using various community detection

945 algorithms on the performance of the proposed community-based methods. It can
 946 be inferred from the results reported in Figure 11, that the performance of the
 947 proposed methods evolves in the same way. Whatever the community strength of
 948 the network, their curves display the same behavior.

949 In networks with a strong community structure (e.g., Power-grid network),
 950 the performance of all the proposed community-based methods for WalkTrap and
 951 Louvain detection algorithms display roughly the same behavior. Their perfor-
 952 mance is, however, slightly better for Infomap algorithm. The gain is around 5%
 953 for all three strategies, as it is reported in Figure 11 (a). In networks with a loose
 954 community structure (e.g., Georgetown network), the performance of the proposed
 955 community-based strategies with Louvain algorithm is worse than when the com-
 956 munity detection is performed using WalkTrap and Infomap algorithms. The gain
 957 is around 10% and 17% when employing WalkTrap and Infomap respectively, as
 958 it is shown in Figure 11 (b). The best results are obtained by using Infomap
 959 algorithm. That being said, the proposed strategies exhibit almost the same per-
 960 formance regardless of the community detection algorithm when the network has
 961 well-separated communities. Their performance is different in networks with an
 962 unclear community structure. It increases when using WalkTrap and Infomap al-
 963 gorithms. Note that the modularity is a good indicator of the performance. Indeed,
 964 the performance increases when the modularity increases.

965 6.3.3 Comparison with the alternative methods

966 Figure 12 shows the relative difference of the outbreak size between the proposed
 967 strategies and the alternative ones as a function of the fraction of the immunized
 968 nodes. The proposed strategies are evaluated on the Power-grid network in (a) and
 969 (b) and the Georgetown network in (c) and (d) for the WalkTrap and Infomap de-
 970 tection algorithms. The left panels of this figure show the comparison between the
 971 Community Hub-bridge and the alternative methods. The middle panels represent
 972 the difference between the Number of Neighboring Communities and the alterna-
 973 tive methods. On the other hand, the right panels show the difference between the
 974 Weighted Community Hub-Bridge method and the alternative ones.

975 In networks with a strong community structure, the performance of the Com-
 976 munity Hub-bridge is still better than the alternative methods with an average
 977 gain of 13% over the best alternative (Community-based Mediator) for Infomap,
 978 while the gain is around 10% for WalkTrap and Louvain algorithms. This method
 979 has a minimal gain of 3% for Infomap. The middle and the right panels of Figure 12
 980 (a) and (b) show that $\Delta R/R_0$ exhibits sometimes a negative value for both the
 981 Number of Neighboring Communities and the Weighted Community Hub-Bridge
 982 strategies. In all the figures, they perform less than CBM and Betweenness as
 983 is the case of Louvain algorithm. Therefore, for all three algorithms, the Com-
 984 munity Hub-Bridge is the most efficient strategy in networks with a well-defined
 985 community structure. Furthermore, it shows its best performance after detecting
 986 communities through Infomap algorithm.

987 In networks with a loose community structure, the Community Hub-Bridge is
 988 performing worse than the Community-Based Betweenness for both WalkTrap and
 989 Infomap, as it is shown in the left panels of Figure 12 (c) and (d). These results
 990 are similar to the ones obtained using Louvain algorithm. On the other hand, the

991 Number of Neighboring Communities outperforms its competitors with an aver-
992 age gain of 3% and 6% over its best alternative (Community-Based Betweenness)
993 for WalkTrap and Infomap respectively (see the middle panels of Figure 12 (c)
994 and (d)), whereas it performs worse than CBB for Louvain. Indeed, the WalkTrap
995 and Infomap algorithms detect a higher number of communities as compared to
996 Louvain algorithm, which uncovers large communities. This makes the ranking
997 using the Number of Neighboring Communities more efficient in the case of Walk-
998 Trap and Infomap. The right panels of Figure 12 (c) and (d) show that $\Delta R/R_0$
999 (taking the Weighted Community Hub-bridge as a reference strategy) always has
1000 a positive value. Its performance is better than the alternative methods with an
1001 average gain of 10% and 15% over the best one (CBB) for WalkTrap and Infomap
1002 respectively, while the gain is around 6% for Louvain. Therefore, the Weighted
1003 Community Hub-Bridge is the most efficient strategy in networks with an unclear
1004 community structure for all the tested detection algorithms. Yet, as expected its
1005 best performance is achieved with Infomap algorithm.

1006 7 Conclusion

1007 The adoption of an appropriate immunization strategy has aroused much interest
1008 among researchers aiming to control any threat of infectious diseases spreading.
1009 Despite the presence of a community structure in all social networks, this property
1010 has been mostly ignored by the existing immunization strategies. In this paper,
1011 three community-based strategies are proposed. They engage more topological
1012 information related to networks with a non-overlapping community structure. The
1013 proposed strategies are evaluated in different synthetic and real networks. To verify
1014 their effectiveness, the SIR epidemic model is employed. First of all, results show
1015 that local strategies underperform compared to global strategies. Indeed, as they
1016 do not have access to the whole network structure, it is not easy to exploit their
1017 properties.

1018 Extensive investigation also shows that generally, the proposed immunization
1019 strategies have a smaller epidemic size compared to the most influential global
1020 immunization strategies (Community-based Mediator and Community-based Be-
1021 tweenness) and the Comm strategy designed for networks with non-overlapping
1022 community structure. The Community Hub-Bridge method is particularly suited
1023 to networks with a strong community structure. The Number of Neighboring
1024 communities shows its best with medium strength community structure while
1025 Weighted Community Hub-Bridge is more efficient in networks with weak com-
1026 munity structure. Additionally, community size range plays an important role in
1027 the diffusion process. Immunization strategies are more efficient when community
1028 size is small. Results to the SIR parameters model variations, show that the immu-
1029 nization strategies display the same type of performances. However, by increasing
1030 the infection rate, a greater proportion of immunized nodes is needed to mitigate
1031 the spread of the epidemic. To test the effectiveness of the proposed strategies
1032 regardless of the epidemiological models, we compute the size of the Largest Con-
1033 nected Component *LCC*. Results show that the proposed methods are still more
1034 efficient than the alternative ones. Moreover, we report also a set of experiments
1035 using the Walktrap and Infomap detection algorithms to uncover communities.
1036 Results of the investigations show that the performance of the proposed methods

exhibits the same behavior in networks with a well defined community structure, this is for all the three community detection algorithms. Their performance is different in networks with an unclear community detection. In this case, the best results are obtained through the Infomap algorithm.

One of the main benefits of this work is to show that significant gains can be achieved by making a better use of the knowledge of the community structure organization. It can be extended in multiple directions. Firstly, these measures can be improved by using finer weights so as to make them more robust to variations in community structure. Now that the impact of community structure strength has been clearly identified, local versions of the proposed strategies need to be designed. Finally, extension to non-overlapping community structures can be considered.

References

1. Porter, M.A., Onnela, J.-P., Mucha, P.J.: Communities in networks. *Notices of the AMS* **56**(9), 1082–1097 (2009)
2. Ferrara, E.: Community structure discovery in facebook. *International Journal of Social Network Mining* **1**(1), 67–90 (2012)
3. Jebabli, M., Cherifi, H., Cherifi, C., Hamouda, A.: User and group networks on youtube: A comparative analysis. In: *Computer Systems and Applications (AICCSA)*, 2015 IEEE/ACS 12th International Conference Of, pp. 1–8 (2015). IEEE
4. Jebabli, M., Cherifi, H., Cherifi, C., Hammouda, A.: Overlapping community structure in co-authorship networks: A case study. In: *U-and e-Service, Science and Technology (UNESST)*, 2014 7th International Conference On, pp. 26–29 (2014). IEEE
5. Liu, Z., Hu, B.: Epidemic spreading in community networks. *EPL (Europhysics Letters)* **72**(2), 315 (2005)
6. Gupta, N., Singh, A., Cherifi, H.: Community-based immunization strategies for epidemic control. In: *Communication Systems and Networks (COMSNETS)*, 2015 7th International Conference On, pp. 1–6 (2015). IEEE
7. Ghalmane, Z., El Hassouni, M., Cherifi, H.: Betweenness centrality for networks with non-overlapping community structure. In: *2018 IEEE Workshop on Complexity in Engineering (COMPENG)*, pp. 1–5 (2018). IEEE
8. Chakraborty, D., Singh, A., Cherifi, H.: Immunization strategies based on the overlapping nodes in networks with community structure. In: *International Conference on Computational Social Networks*, pp. 62–73 (2016). Springer
9. Hurley, M., Jacobs, G., Gilbert, M.: The basic si model. *New Directions for Teaching and Learning* **2006**(106), 11–22 (2006)
10. Newman, M.E.: Spread of epidemic disease on networks. *Physical review E* **66**(1), 016128 (2002)
11. Moreno, Y., Pastor-Satorras, R., Vespignani, A.: Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B-Condensed Matter and Complex Systems* **26**(4), 521–529 (2002)
12. Giabbanelli, P.J., Peters, J.G.: *Reseaux complexes et epidemies*. *TSI-Technique et Science Informatiques* **30**(2), 181 (2011)
13. Cohen, R., Havlin, S., Ben-Avraham, D.: Efficient immunization strategies for computer networks and populations. *Physical review letters* **91**(24), 247901 (2003)
14. Bomze, I.M., Budinich, M., Pardalos, P.M., Pelillo, M.: The maximum clique problem. In: *Handbook of Combinatorial Optimization*, pp. 1–74. Springer, ??? (1999)
15. Fortunato, S.: Community detection in graphs. *Physics reports* **486**(3-5), 75–174 (2010)
16. Palla, G., Derényi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. *nature* **435**(7043), 814 (2005)
17. Leskovec, J., Lang, K.J., Dasgupta, A., Mahoney, M.W.: Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics* **6**(1), 29–123 (2009)
18. Jeub, L.G., Balachandran, P., Porter, M.A., Mucha, P.J., Mahoney, M.W.: Think locally, act locally: Detection of small, medium-sized, and large communities in large networks. *Physical Review E* **91**(1), 012821 (2015)

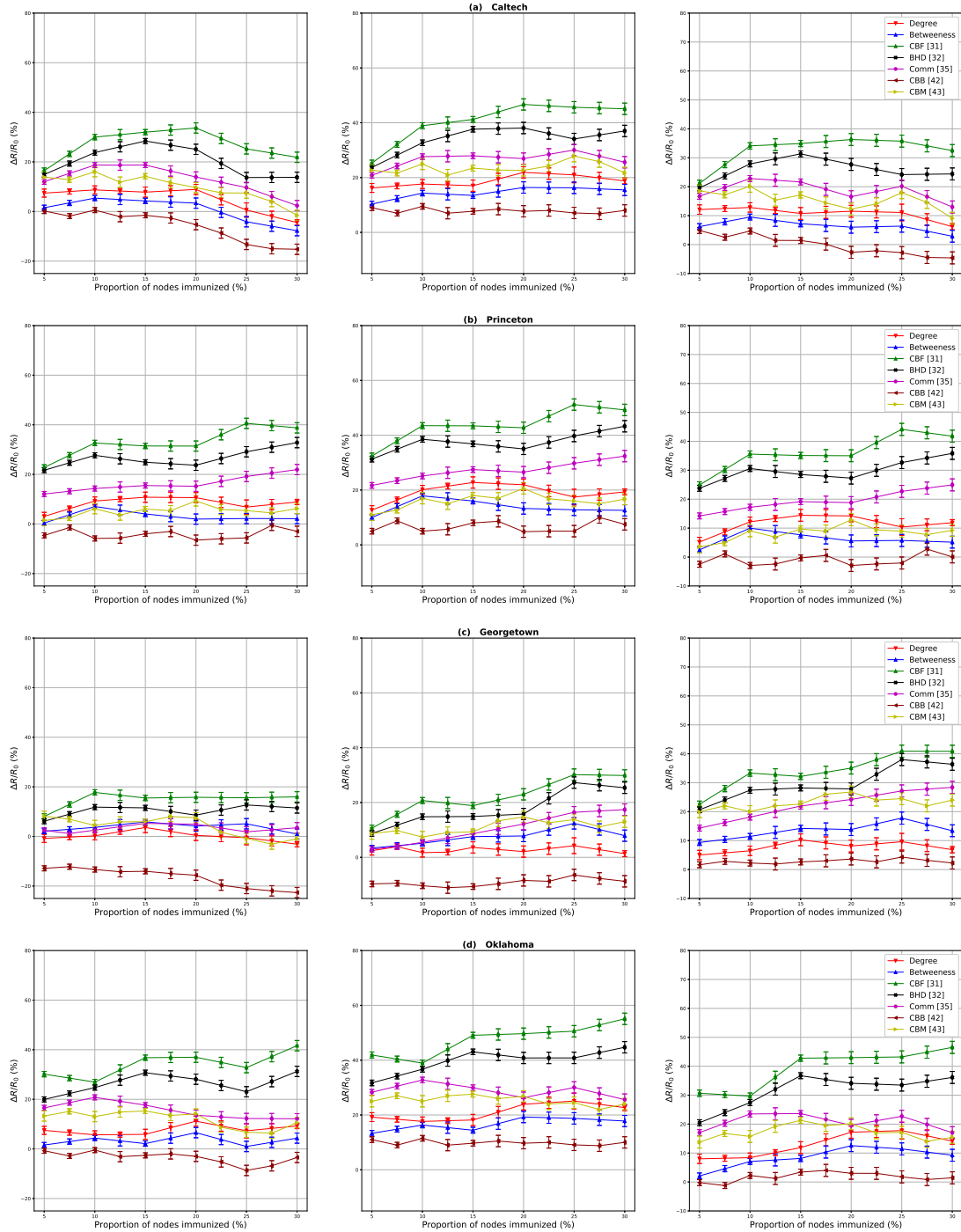
- 1092 19. Yang, J., Leskovec, J.: Structure and overlaps of ground-truth communities in networks.
1093 ACM Transactions on Intelligent Systems and Technology (TIST) **5**(2), 26 (2014)
- 1094 20. Newman, M.E.: Modularity and community structure in networks. Proceedings of the
1095 national academy of sciences **103**(23), 8577–8582 (2006)
- 1096 21. Zhang, X.-S., Wang, R.-S., Wang, Y., Wang, J., Qiu, Y., Wang, L., Chen, L.: Modularity
1097 optimization in community detection of complex networks. EPL (Europhysics Letters)
1098 **87**(3), 38002 (2009)
- 1099 22. Newman, M.E.: Fast algorithm for detecting community structure in networks. Physical
1100 review E **69**(6), 066133 (2004)
- 1101 23. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of commu-
1102 nities in large networks. Journal of statistical mechanics: theory and experiment **2008**(10),
1103 10008 (2008)
- 1104 24. Pons, P., Latapy, M.: Computing communities in large networks using random walks. J.
1105 Graph Algorithms Appl. **10**(2), 191–218 (2006)
- 1106 25. Rosvall, M., Bergstrom, C.T.: An information-theoretic framework for resolving commu-
1107 nity structure in complex networks. Proceedings of the National Academy of Sciences
1108 **104**(18), 7327–7331 (2007)
- 1109 26. Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal com-
1110 munity structure. Proceedings of the National Academy of Sciences **105**(4), 1118–1123
1111 (2008)
- 1112 27. Fortunato, S., Hric, D.: Community detection in networks: A user guide. Physics Reports
1113 **659**, 1–44 (2016)
- 1114 28. Orman, K., Labatut, V., Cherifi, H.: An empirical study of the relation between community
1115 structure and transitivity. In: Complex Networks, pp. 99–110 (2013)
- 1116 29. Ghalmane, Z., El Hassouni, M., Cherifi, C., Cherifi, H.: k-truss decomposition for modular
1117 centrality. In: 2018 9th International Symposium on Signal, Image, Video and Communi-
1118 cations (ISIVC), pp. 241–248 (2018). IEEE
- 1119 30. Ghalmane, Z., El Hassouni, M., Cherifi, C., Cherifi, H.: Centrality in modular networks.
1120 EPJ Data Science **8**(1), 15 (2019)
- 1121 31. Salathé, M., Jones, J.H.: Dynamics and control of diseases in networks with community
1122 structure. PLoS computational biology **6**(4), 1000736 (2010)
- 1123 32. Gong, K., Tang, M., Hui, P.M., Zhang, H.F., Younghae, D., Lai, Y.-C.: An efficient im-
1124 munization strategy for community networks. PloS one **8**(12), 83489 (2013)
- 1125 33. Taghavian, F., Salehi, M., Teimouri, M.: A local immunization strategy for networks with
1126 overlapping community structure. Physica A: Statistical Mechanics and its Applications
1127 **467**, 148–156 (2017)
- 1128 34. Lü, L., Chen, D., Ren, X.-L., Zhang, Q.-M., Zhang, Y.-C., Zhou, T.: Vital nodes identifi-
1129 cation in complex networks. Physics Reports **650**, 1–63 (2016)
- 1130 35. Gupta, N., Singh, A., Cherifi, H.: Centrality measures for networks with community struc-
1131 ture. Physica A: Statistical Mechanics and its Applications **452**, 46–59 (2016)
- 1132 36. Barabási, A.-L., Bonabeau, E.: Scale-free networks. Scientific american **288**(5), 60–69
1133 (2003)
- 1134 37. Ferrara, E., De Meo, P., Fiumara, G., Provetti, A.: The role of strong and weak ties in face-
1135 book: a community structure perspective. Preprint at <http://arXiv.org/abs/1203.0535>
1136 (2012)
- 1137 38. Hébert-Dufresne, L., Allard, A., Young, J.-G., Dubé, L.J.: Global efficiency of local im-
1138 munization on complex networks. Scientific reports **3**, 2171 (2013)
- 1139 39. Kumar, M., Singh, A., Cherifi, H.: An efficient immunization strategy using overlapping
1140 nodes and its neighborhoods. In: Companion of the The Web Conference 2018 on The
1141 Web Conference 2018, pp. 1269–1275 (2018). International World Wide Web Conferences
1142 Steering Committee
- 1143 40. Newman, M.E.: A measure of betweenness centrality based on random walks. Social net-
1144 works **27**(1), 39–54 (2005)
- 1145 41. Brandes, U.: A faster algorithm for betweenness centrality. Journal of mathematical soci-
1146 ology **25**(2), 163–177 (2001)
- 1147 42. Kitromilidis, M., Evans, T.S.: Community detection with metadata in a network of bi-
1148 ographies of western art painters. arXiv preprint arXiv:1802.07985 (2018)
- 1149 43. Tulu, M.M., Hou, R., Younas, T.: Identifying influential nodes based on community struc-
1150 ture to speed up the dissemination of information in complex network. IEEE ACCESS **6**,
1151 7390–7401 (2018)

- 1152 44. Lancichinetti, A., Fortunato, S., Radicchi, F.: Benchmark graphs for testing community
1153 detection algorithms. *Physical review E* **78**(4), 046110 (2008)
- 1154 45. Orman, G.K., Labatut, V., Cherifi, H.: Towards realistic artificial benchmark for commu-
1155 nity detection algorithms evaluation. *International Journal of Web Based Communities*
1156 **9**(3), 349–370 (2013)
- 1157 46. Albert, R., Barabási, A.-L.: Statistical mechanics of complex networks. *Reviews of modern*
1158 *physics* **74**(1), 47 (2002)
- 1159 47. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.-U.: Complex networks:
1160 Structure and dynamics. *Physics reports* **424**(4-5), 175–308 (2006)
- 1161 48. Newman, M.E.: The structure and function of complex networks. *SIAM review* **45**(2),
1162 167–256 (2003)
- 1163 49. Traud, A.L., Mucha, P.J., Porter, M.A.: Social structure of facebook networks. *Physica A:*
1164 *Statistical Mechanics and its Applications* **391**(16), 4165–4180 (2012)
- 1165 50. Watts, D.J., Strogatz, S.H.: Collective dynamics of small-world networks. *nature*
1166 **393**(6684), 440 (1998)
- 1167 51. Leskovec, J., Kleinberg, J., Faloutsos, C.: Graph evolution: Densification and shrinking
1168 diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)* **1**(1), 2 (2007)
- 1169 52. Orman, G.K., Labatut, V., Cherifi, H.: On accuracy of community structure discovery
1170 algorithms. *arXiv preprint arXiv:1112.4134* (2011)
- 1171 53. Orman, G.K., Labatut, V., Cherifi, H.: Comparative evaluation of community detection
1172 algorithms: a topological approach. *Journal of Statistical Mechanics: Theory and Experiment*
1173 **2012**(08), 08001 (2012)
- 1174 54. Wang, W., Liu, Q.-H., Zhong, L.-F., Tang, M., Gao, H., Stanley, H.E.: Predicting the
1175 epidemic threshold of the susceptible-infected-recovered model. *Scientific reports* **6**, 24676
1176 (2016)
- 1177 55. Danon, L., Diaz-Guilera, A., Duch, J., Arenas, A.: Comparing community structure identi-
1178 fication. *Journal of Statistical Mechanics: Theory and Experiment* **2005**(09), 09008 (2005)

1179 Additional material

Table 6 The Analysis of Variance ANOVA with a significance level $\alpha = 0.05$. SS is the sum of squares. df is the degree of freedom. MS is the mean square. F is the test statistic. P -value is the probability value and F_{critic} is the critical value of F.

	Source of variations	SS	df	MS	F	P -value	F_{crit}
Comm strategy							
$\mu = 0.1$; [100,500]	Between groups	4.256503573	1	4.256503573	0.019440919	0.890504334	4.351243478
$\mu = 0.1$; [50,250]	Within groups	4378.911918	20	218.9455959			
$\mu = 0.4$; [100,500]	Between groups	3.25263213	1	3.25263213	0.022543556	0.882154017	4.351243478
$\mu = 0.4$; [50,250]	Within groups	2885.642407	20	144.2821203			
$\mu = 0.7$; [100,500]	Between groups	14.15562603	1	14.15562603	0.050262847	0.824880951	4.351243478
$\mu = 0.7$; [50,250]	Within groups	5632.639949	20	281.6319974			
$\mu = 0.9$; [100,500]	Between groups	1.420498127	1	1.420498127	0.00563914	0.967814872	4.351243478
$\mu = 0.9$; [50,250]	Within groups	5037.9883	20	251.899415			
CBB strategy							
$\mu = 0.1$; [100,500]	Between groups	2.968469205	1	2.968469205	0.007615194	0.931328629	4.351243478
$\mu = 0.1$; [50,250]	Within groups	7796.175363	20	389.8087681			
$\mu = 0.4$; [100,500]	Between groups	1.419799471	1	1.419799471	0.013046107	0.910202616	4.351243478
$\mu = 0.4$; [50,250]	Within groups	2176.587136	20	108.8293568			
$\mu = 0.7$; [100,500]	Between groups	2.78891	1	2.78891	0.01472556	0.999564357	4.351243478
$\mu = 0.7$; [50,250]	Within groups	3787.848287	20	189.3924143			
$\mu = 0.9$; [100,500]	Between groups	12.76013782	1	12.76013782	0.03371355	0.856166211	4.351243478
$\mu = 0.9$; [50,250]	Within groups	7569.738518	20	378.4869259			
CBM strategy							
$\mu = 0.1$; [100,500]	Between groups	6.632987506	1	6.632987506	0.037059628	0.849285018	4.351243478
$\mu = 0.1$; [50,250]	Within groups	3579.629836	20	178.9814918			
$\mu = 0.4$; [100,500]	Between groups	2.153661701	1	2.153661701	0.01532581	0.973948282	4.351243478
$\mu = 0.4$; [50,250]	Within groups	2810.5024	20	140.52512			
$\mu = 0.7$; [100,500]	Between groups	1.349802017	1	1.349802017	0.005821325	0.939940411	4.351243478
$\mu = 0.7$; [50,250]	Within groups	4637.43895	20	231.8719475			
$\mu = 0.9$; [100,500]	Between groups	0.54374064	1	0.54374064	0.003027532	0.956666083	4.351243478
$\mu = 0.9$; [50,250]	Within groups	3591.973201	20	179.59866			



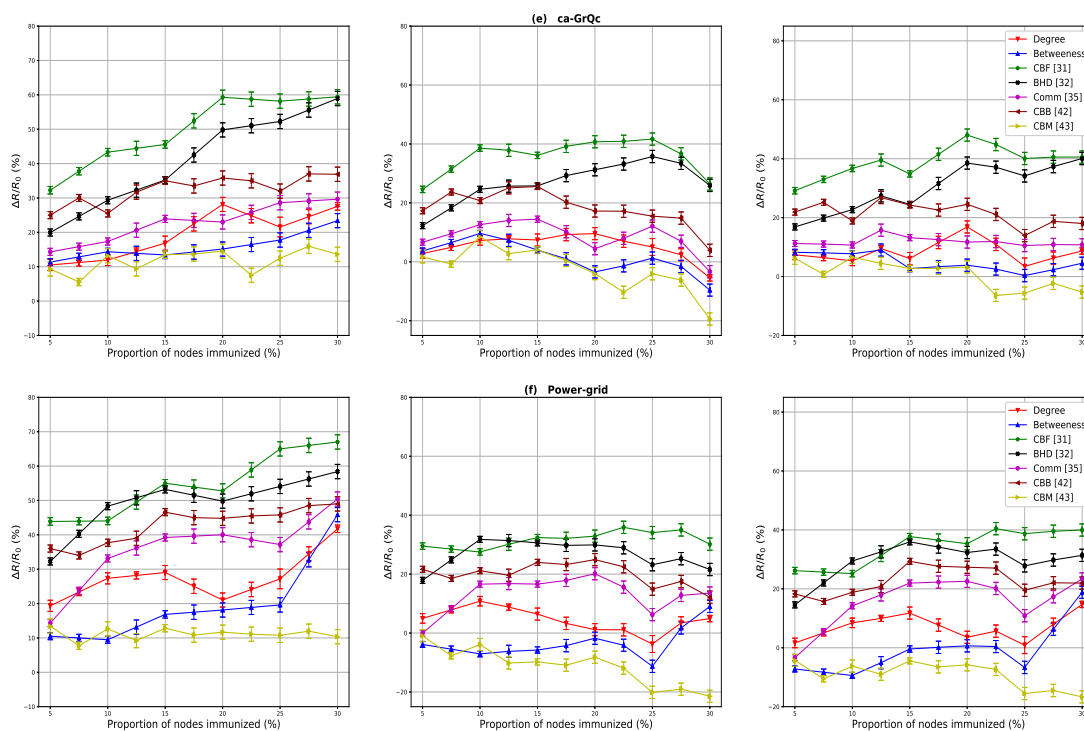


Fig. 8 The relative difference of the outbreak size $\Delta R/R_0$ as a function of the fraction of immunized nodes. The left panels show the difference between Community Hub-Bridge method and the alternative methods. The middle and the right panels show respectively the difference between the Number of Neighboring Communities method, the Weighted Community Hub-Bridge and the alternative methods. We note that a positive value of $\Delta R/R_0$ means a higher performance of the proposed method. Simulations are performed on different types of real-world networks. Final values are obtained by running 600 independent simulations per network, immunization coverage and immunization method.

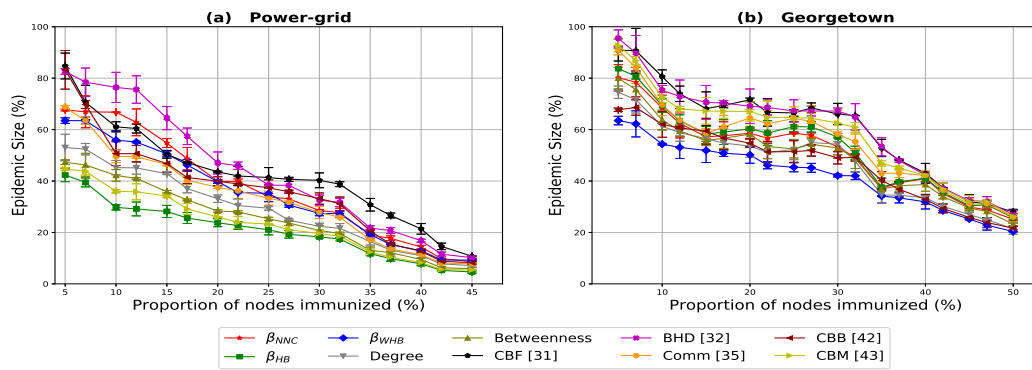


Fig. 9 Effect of different immunization methods on the epidemic size during the SIR simulations performed on (a) Power-grid and (b) Georgetown network, with $\lambda = 0.9$ and $\gamma = 0.2$.

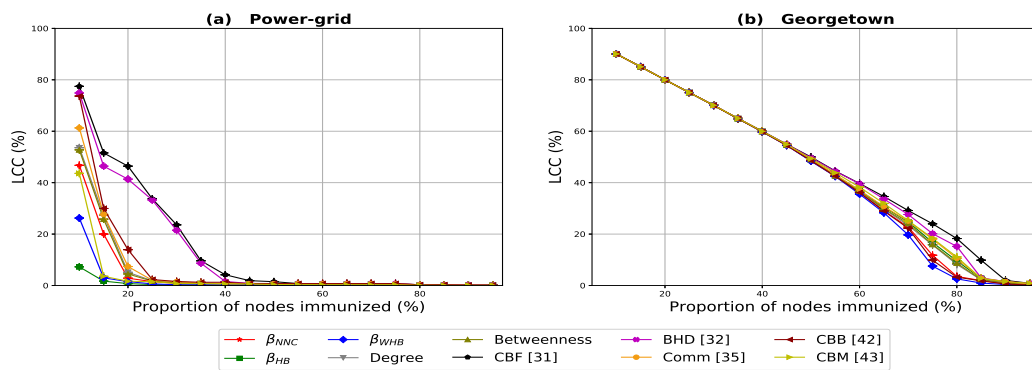


Fig. 10 The size of the Largest Connected Component (LCC) for various immunization strategies performed on (a) Power-grid and (b) Georgetown network. Each point is the result of the LCC size as function of the proportion of the immunized nodes.

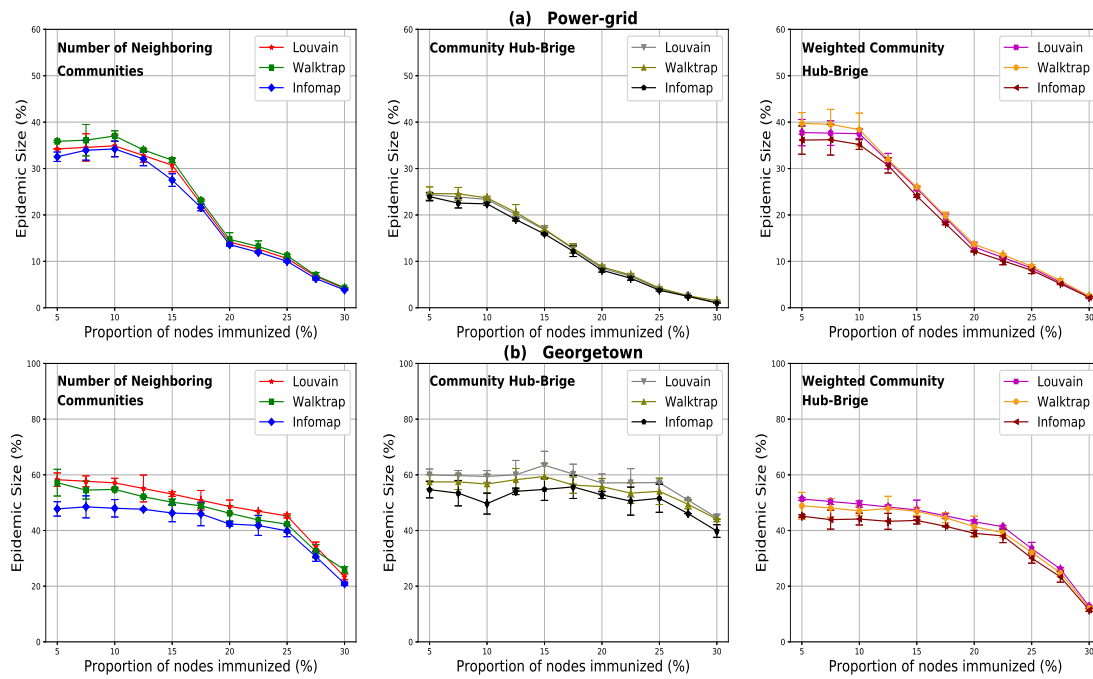


Fig. 11 Effect of various community detection algorithms on the performance of the proposed community based methods. Each point shows the epidemic size with respect to the proportion of the immunized nodes.

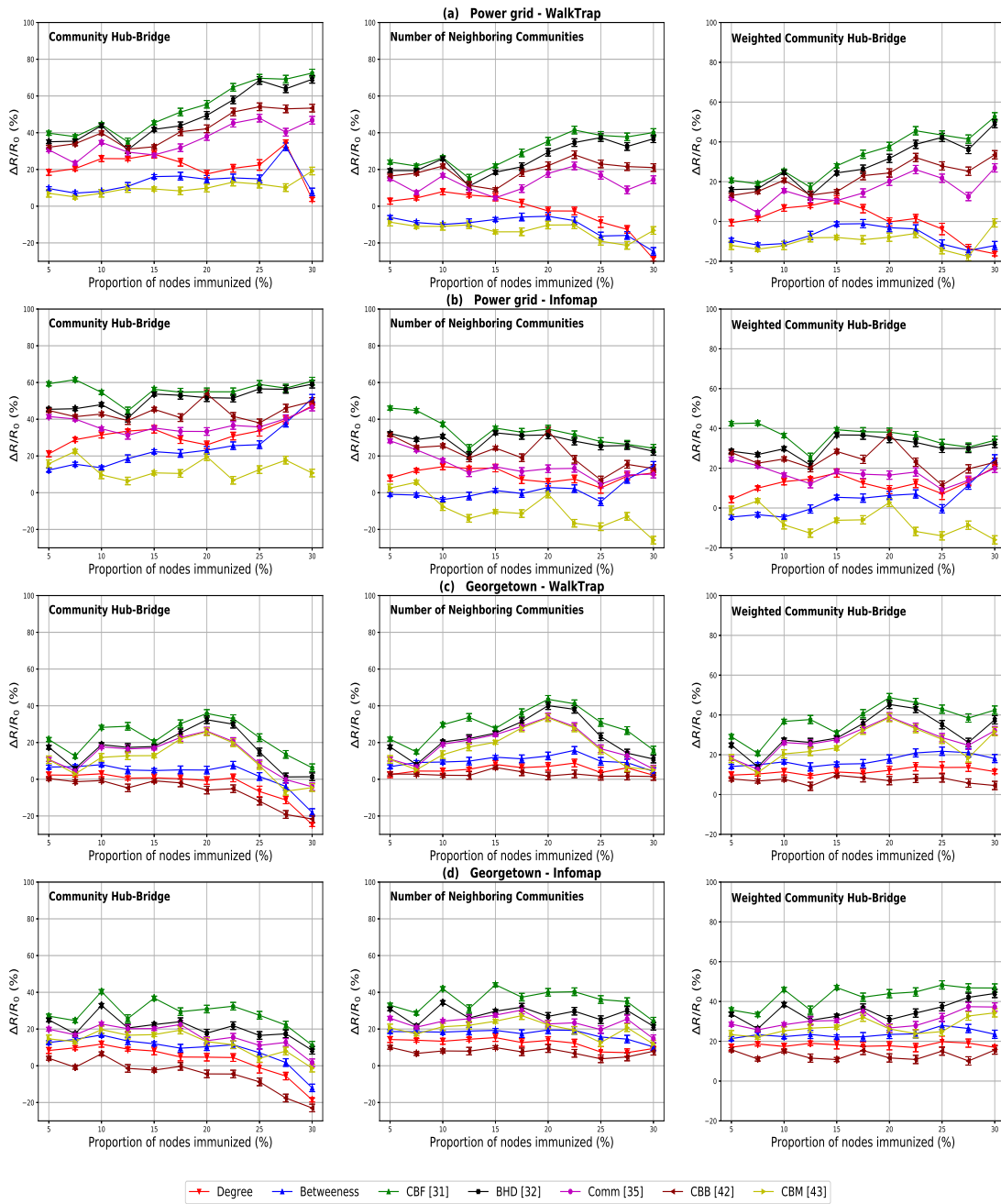


Fig. 12 The relative difference of the outbreak size as a function of the proportion of the immunized nodes. The left panels show the difference between Community Hub-Bridge method and the alternative methods. The middle panels show the difference between the Number of Neighboring Communities method and the alternative methods. The right panels show the difference between the Weighted Community Hub-Bridge and the alternative methods. The immunization methods are performed on Power grid network in (a) and (b) and Georgetown network in (c) and (d) for the WalkTrap and Infomap algorithms.